University of Denver

Digital Commons @ DU

6-2023

Patient Movement Monitoring Based on IMU and Deep Learning

Mohsen Sharifi Renani
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Artificial Intelligence and Robotics Commons, Biomechanical Engineering Commons, Biomechanics and Biotransport Commons, and the Biomedical Devices and Instrumentation Commons

Patient Movement Monitoring Based on IMU and Deep Learning

Abstract

Osteoarthritis (OA) is the leading cause of disability among the aging population in the United States and is frequently treated by replacing deteriorated joints with metal and plastic components. Developing better quantitative measures of movement quality to track patients longitudinally in their own homes would enable personalized treatment plans and hasten the advancement of promising new interventions. Wearable sensors and machine learning used to quantify patient movement could revolutionize the diagnosis and treatment of movement disorders. The purpose of this dissertation was to overcome technical challenges associated with the use of wearable sensors, specifically Inertial Measurement Units (IMUs), as a diagnostic tool for osteoarthritic (OA) and total knee replacement patients (TKR) through a detailed biomechanical assessment and development of machine learning algorithms. Specifically, the first study developed a relevant dataset consisting of IMU and associated biomechanical parameters of OA and TKR patients performing various activities, created a machine learning-based framework to accurately estimate spatiotemporal movement characteristics from IMU during level ground walking, and defined optimum sensor configuration associated with the patient population and activity. The second study designed a framework to generate synthetic kinematic and associated IMU data as well as investigated the influence of adding synthetic data into training-measured data on deep learning model performance. The third study investigated the kinematic variation between two patient's population across various activities: stair ascent, stair descent, and gait using principle component analysis PCA. Additionally, PCA-based autoencoders were developed to generate synthetic kinematics data for each patient population and activity. The fourth study investigated the potential use of a universal deep learning model for the estimation of lower extremities' kinematics across various activities. Therefore, this model can be used as a global model for transfer learning methods in future research. This line of study resulted in a machine-learning framework that can be used to estimate biomechanical movements based on a stream of signals emitted from low-cost and portable IMUs. Eventually, this could lead to a simple clinical tool for tracking patients' movements in their own homes and translating those movements into diagnostic metrics that clinicians will be able to use to tailor treatment to each patient's needs in the future.

Sciences | Engineering

Publication Statement

Patient Movement Monitoring Based on IMU And Deep Learning

_____

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

_____

by

Mohsen Sharifi Renani

June 2023

Advisor: Chadd W. Clary

Author: Mohsen Sharifi Renani
Title: Patient Movement Monitoring Based on IMU And Deep Learning
Advisor: Chadd W. Clary
Degree Date: June 2023

ABSTRACT

Osteoarthritis (OA) is the leading cause of disability among the aging population in the United States and is frequently treated by replacing deteriorated joints with metal and plastic components. Developing better quantitative measures of movement quality to track patients longitudinally in their own homes would enable personalized treatment plans and hasten the advancement of promising new interventions. Wearable sensors and machine learning used to quantify patient movement could revolutionize the diagnosis and treatment of movement disorders. The purpose of this dissertation was to overcome technical challenges associated with the use of wearable sensors, specifically Inertial Measurement Units (IMUs), as a diagnostic tool for osteoarthritic (OA) and total knee replacement patients (TKR) through a detailed biomechanical assessment and development of machine learning algorithms. Specifically, the first study developed a relevant dataset consisting of IMU and associated biomechanical parameters of OA and TKR patients performing various activities, created a machine learning-based framework to accurately estimate spatiotemporal movement characteristics from IMU during level ground walking, and defined optimum sensor configuration associated with the patient population and activity. The second study designed a framework to generate synthetic kinematic and associated IMU data as well as investigated the influence of adding synthetic data into training-

measured data on deep learning model performance. The third study investigated the kinematic variation between two patient's population across various activities: stair ascent, stair descent, and gait using principle component analysis PCA. Additionally, PCA-based autoencoders were developed to generate synthetic kinematics data for each patient population and activity. The fourth study investigated the potential use of a universal deep learning model for the estimation of lower extremities' kinematics across various activities. Therefore, this model can be used as a global model for transfer learning methods in future research. This line of study resulted in a machine-learning framework that can be used to estimate biomechanical movements based on a stream of signals emitted from low-cost and portable IMUs. Eventually, this could lead to a simple clinical tool for tracking patients' movements in their own homes and translating those movements into diagnostic metrics that clinicians will be able to use to tailor treatment to each patient's needs in the future.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Introductions

Human movement analysis aims at gathering quantitative information about the mechanics during the execution of a motor task (Cappozzo et al. 2005) and has a long history, dating back to when the ancient Greeks (500-300 BC) describe human movement (Andriacchi and Alexander 2000; Nigg and Herzog 2007). During this period, scientists also developed basic abilities (such as mathematics, mechanics, and medicine) to better describe movement and to develop underlying theories (Nigg and Herzog 2007). Later on, during the scientific revolution, Newton's laws were established, which provided the impetus for the study of human movement and the tools to understand it (Nigg and Herzog 2007).

In recent decades, human movement analysis has turned into an open area for researchers due to the advancement of technology and expanding the broad spectrum of its application such as athletic performance, surveillance identification, disease diagnostics, monitoring, device design, rehabilitation, and entertainment (e.g. metaverse) (Singh et al. 2018). More specifically in the orthopedic domain, developing better quantitative metrics of movement quality to track osteoarthritis (OA) and Joint Replacement (JR) patients in a wide range of environments is very important. It would allow clinicians to monitor the

progression of Osteoarthritis, advance personalized treatment plans, optimize timing for joint replacement, and develop more effective post-operative rehabilitation plans.

We can categorize current movement analysis technologies into three groups. The first group is a non-contact or vision-based system such as a markerless motion capture system, video camera, or Kinect. The main limitation of these devices is being constrained to specific environments. The second group is the conventional marker-based motion capture systems. These systems usually consist of reflective markers which attached to the body and multiple cameras for tracking those markers, thereby called hybrid methods which is benefiting from both contact and non-contact-based technologies. Despite its reliability and being considered a gold standard, using this technology is extremely time-consuming, limited to the lab environment, expensive, and requires technical expertise. The last group is contact-based devices which use the transmitted signals to evaluate movements. Inertia measurement units (IMUs) and electromagnetic units are two examples of these technologies. Unlike other forms of motion capture technology, IMU devices can be worn in any environment, they are inexpensive and accessible, and they can track movement throughout time. However, IMUs have some limitations, such as sensor artifact, challenges in interpretation, and required device-to-body attachment.

An IMU consists of a three-axis accelerometer and a three-axis gyroscope, making it a six-axis unit. An additional 3-axis magnetometer can also be included, making it a 9-axis IMU. An IMU can measure acceleration, angular velocity, and magnetic fields, as well as determine motion, orientation, and heading when paired with sensor fusion techniques.

The IMUs are cost-effective and can evaluate patient movement throughout the time and in any environment. However, they suffer from an enormous challenge which is interpreting the IMU's noisy data into clinically meaningful metrics. Advanced machine learning models can overcome this challenge and improve the diagnosis and treatment of movement disorders by providing a quantitative assessment of movements such as activity type, spatial-temporal parameters, joint range of motion, and patient outcome.

1.2 Objectives

The overall goal of this study is to develop a tool to measure human movement using IMUs and advanced machine learning (ML) techniques. Four specific objectives were proposed to accomplish this goal. The first specific aim was to develop an ML model to classify the human activities of daily living based on IMU signals. The second specific objective was to provide a comprehensive measurement of spatial-temporal parameters of the most common movement of gait using deep learning and IMUs for patients with OA and total knee arthroplasty (TKA). The effect of various IMU sensor placements on the measurement of gait parameters was also investigated. The third specific objective was to develop a pipeline to create synesthetic IMU and use it for the training of the DL model to estimate joint kinematics. The fourth objective was to use principal component analysis methods to investigate the joint kinematics of OA and TKA patients across gait, stair ascent, and strait descent. Additionally, PCA was leveraged to develop a PCA-based auto-encoder to generate synthetic kinematics data. The last objective of this study was to develop a single multi-purpose deep learning model based on transformer architecture,

BioMAT, for the prediction of joint kinematic from IMUs across activities of daily living such as level ground, stair ascent/descent, and ramp ascent/descent.

1.3 Dissertation Overview

Chapter 2 provides a review of recent literature associated with the field of human movement analysis using machine learning.

Chapter 3 presents *Deep Learning in Gait Parameter Prediction for OA and TKA Patients Wearing IMU* whose objective was to evaluate the feasibility of using deep learning methods in estimating spatial-temporal gait parameters for OA and TKA patients. The performance of various neural netweork architectures was assessed for estimating the stride length and the optimal model was used for training and evaluating the other 11 spatial-temporal parameters. The optimum sensor configuration was also identified. This study has been published in the journal *Sensors* (M. Sharifi Renani et al., 2020).

Chapter 4 presents *The Use of Synthetic IMU Signals in the Training of Deep Learning Models Significantly Improves the Accuracy of Joint Kinematic Predictions* whose objective was to evaluate the feasibility of using deep learning methods as well as effectiveness of synthetic kinematic and imu data in estimating hip and knee joint kinematics during gait activities for OA and TKA patients. This study has been published in the journal *Sensors* (M. Sharifi Renani et al., 2021).

Chapter 5 presents *Principal component analysis of OA and TKA patients across the activities of daily living: Gait, Stair Ascent, Stair Descent, Sit to Stand* whose objective was to identify the different modes of kinematic variation between OA and TKA in different activities; second, to identify the relationship between mode variations of

different activities. Finally, to investigate the use of PCA for generating synthetic kinematic data generation, which can be employed for future ML applications and musculoskeletal analysis. This study has been submitted to the *Journal of Gait and Posture.*

Chapter 6 presents *BioMAT: An Open-Source Biomechanics Multi-Activities Transformer (MAT) for Joint Kinematic Prediction based on IMUs* whose objective was to propose a single open-source deep learning model based on an attention mechanism and transform models to estimate joint kinematics across various activities. This study will be submitted to the journal *Sensors.*

Chapter 7 summarizes the contributions of the studies presented in this dissertation and suggests directions for future work in this field.

CHAPTER 2: BACKGROUND INFORMATION AND LITERATURE REVIEW

The phrase "human motion analysis" refers to any method for obtaining a quantitative or qualitative measurement of human motion (Inman et al. 1981). Quantitative analysis involves the measurement of biomechanical variables, such as movement of the whole-body center of mass, the relative movement between adjacent bones, or joint kinematics; the forces exchanged with the environment; the resultant loads transmitted across sections of body segments or between body segments, or transmitted by individual body tissues such as muscles, tendons, ligaments, and bones; and body segment energy variation and muscular work (Cappozzo et al. 2005). Researchers and specialists can quantitatively assess the motion parameters of patients through human motion analysis. Accurate measurements of body movements are crucial to the identification of biomechanical disorders, abnormal neuromuscular control, and the prevention of injuries (Klette and Tee 2008).

2.1 Tools

2.1.1 Motion Capture Systems

Various systems have been developed and used to analyze human movement. In controlled environments, specialized systems like Vicon (Vicon Motion Systems Ltd., Oxford, UK), a motion capture system (MOCAP), and Optotrak (Northern Digital Inc.,

Ontario, Canada) are highly accurate. They used a set of cameras calibrated and correlated in a specific location to record the position of retro-reflective markers or infrared emitting diodes attached to the body (Lopez-Nava and Munoz-Melendez 2016). These systems are considered the laboratory's gold standard and are used to evaluate the performance of other new systems. Despite their high accuracy, they are expensive, limited to the laboratory environment, require expertise to set up and analyze data, and are time-consuming which makes them undesirable for large-scale population studies. The majority of biomechanics studies in this field use similar tools. In recent years, there has been an increase in demand for the analysis of human movement as a tool for surgical decision-making, diagnosis and monitoring of diseases like Parkinson, rehabilitation in clinics and at home, as well as better understanding sports injuries and recovery. In response to this increased demand and the limitations of the current method for capturing human movement, researchers have been researching alternative methods (Mündermann et al. 2006).

2.1.2 Depth Camera (Kinect)

Other vision-based methods include ambulatory systems, such as those using a Kinect (Microsoft Corporation, WA, USA) to capture human motion. In contrast to MOCAP, these systems are low-cost, portable, and easy to set up. However, they operate in relatively uncontrollable environments with a limited field of view. Additionally, they are intended primarily for indoor use and have limited maneuverability (Chen et al. 2013). Studies have used the Kinect to collect and analyze human movement, for spatial-temporal parameters (Eltoukhy et al. 2016; Mentiplay et al. 2015) and kinematic measures (Oh et al. 2018; Usami et al. 2022; Yamamoto et al. 2021), while validating against the current gold

standard. According to Clark et al. (Clark et al. 2013), gait speed, step length, and stride length showed excellent agreement with 3D marker-based gait analysis with r-values of 0.95, 0.99, and 0.99 respectively and p-values of 0.001. Based on the results of Geerse et al.,(Geerse et al. 2015), a reasonable agreement was found for gait speed, cadence, step length, step time, and stride time, with ICC values of 0.995, 0.974, 0.994, 0.999, 0.888, and 0.962, respectively. The systemic review showed a larger variation in the results for kinematic measures between the Kinect and the current gold standard. Studies showed some kinematic measures with excellent agreement while some kinematic measures presented low agreement and large error. Skals et al. (Skals et al. 2017) found the highest correlations between two methods for knee flexion/extension (0.81), hip flexion/extension (0.82), and hip abduction/adduction (0.81), and the lowest correlation for hip internal/esterla rotation (-0.63) and ankle plantar/dorsi flexion (0.57).Despite similar patterns between the two systems, Anderson et al. (Skipper et al. 2013) found that the Kinect predicted larger peak values for hip joint angles.

2.1.3 Biplane Radiography System

The Mobile Biplane X-Ray or radio graphic Imaging System is an advanced medical imaging tool used to measure three-dimensional dynamic joint motion during overground gait. This system consists of two X-ray sources and two detectors that are mounted on mobile platforms, which can be easily moved to different locations to capture multiple angles of joint motion. This biplane imaging system offers superior accuracy in visualizing the joint from two perpendicular planes, resulting in the most precise measurements of joint position and motion during gait when compared to other movement tracking systems (Gray

8

et al. 2018; Hume et al. 2018). The maximum root-mean-squared errors for translations and rotations of the TKA knee were 0.33 mm and 0.65°, respectively, while for the intact knee, the errors were 0.78 mm and 0.77° for translations and rotations, respectively (Guan et al. 2016).

2.1.4 Wearable/IMUs

The field of motion analysis has recently witnessed significant development thanks to new technology borrowed from aerospace, industrial, and robotic engineering. In the absence of the constraints described above, small, low-power electromechanical sensors using accelerometers, magnetometers, and gyroscopes showed promising results in the provision of dynamic three-dimensional motion analysis. Numerous studies have reported using systems based on different types of inertial sensors, including those based on accelerometers, gyroscopes, magnetometers, or combinations for the study of human motions. The accuracy of human motion measurement relative to the gold standard remains under investigation (Cuesta-Vargas et al. 2010). In a review study, Irvin Hussein Lopes divide the application of human motion analysis using wearable sensors into two classes: one movement measurement and movement classification. The first class is related to measuring or quantifying movements of a specific segment of the human body, such as the limbs or joints, or moments. The outcome of these systems can be a common unit of measurement such as angles, moments, or forces. The second class relates to a high-level classification of human movements, such as "running" or "walking" (Lopez-Nava and Munoz-Melendez 2016).

Biomechanical models can be easily built using MOCAP markers placed on bony landmarks. The data captured on wearable sensors, however, is not always fixed to bony landmarks, and it is measured in a local (i.e. device) reference frame. In the case of combining wearable devices with traditional biomechanical modeling, placement of the wearables may need to be limited to bony regions. In this way, it would be possible to define the device in the coordinate system of particular segments in the model. Then, the complex coordinate transformation would be needed to convert the wearable data from its local coordinate frame to the body's segment and consequently laboratory coordinate system (McCabe 2020). Additionally, raw sensor measurements are prone to noise and non-zero biases because of their microelectromechanical architecture. Furthermore, IMUs are usually not aligned with bone, so it is necessary to determine the misalignment of anatomical coordinate frames and the initial sensor orientation (Weygers et al. 2020). In order to overcome these sensor deficiencies, previous researchers relied on application-specific prior information and assumptions for calculating the joint kinematics or movement analysis (Weygers et al. 2020).

Some of these assumptions were:

1) Assuming the biomechanical system provides reliable a priori information (e.g. range of motion). These assumptions can be used to bound the kinematic solution to a certain range of normal activity. For example, range of motion (ROM) boundaries cannot be generalized across patient populations who may be hypomobile, exceeding the normal range of motion.

2) Assuming joint centers are located at a fixed point. This assumption will be violated by joint translations and soft tissue artifacts between the sensor and the bone (Andriacchi and Alexander 2000; Frick and Rahmatalla 2018b, 2018a)

3) Assuming periodicity in movements to negate integration drift by making the beginning and end of a gait trial equal (Grisetti et al. 2010; Morris 1973)

4) Assuming bilateral symmetry of movement, thus reducing the required the number of sensors on the body. This may over-constrain the system and cannot be applied to pathological patients with asymmetric movements (Bonnet et al. 2013; OHTAKI et al. 2001).

5) Assuming zero-acceleration when the foot is in contact with the ground and expecting one foot to be consistently on the ground during an activity. This assumption can be violated for movements which lack a regular mid-stance phase such as jogging, running, or jumping (Weygers et al. 2020).

2.2 Machine Learning

Machine learning is a field of computer science that uses statistical techniques to enable computers to "learn" with data, without being explicitly programmed. The history of machine learning can be traced back to the early days of artificial intelligence when researchers attempted to create programs that could simulate human learning. However, it was not until the late 1950s and early 1960s that machine learning began to take its modern form. In 1957, Frank Rosenblatt (Rosenblatt 1958) proposed the concept of a neural network, which would become one of the key components in many machine-learning algorithms. In 1959, Arthur Samuel wrote a program called checkers playing program,

which used a simple form of reinforcement learning to improve its performance over time (Samuel 1959). These early efforts laid the foundation for modern machine-learning research.

Machine learning is a subbranch of artificial intelligence and can be categorized into three main groups: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, models learn a mapping between input examples and the target labeled output variables. The problem of supervised learning can be further divided into classification and regression tasks. In the classification task, the model predicts the class label. In regression task, the model predicts the numeral label. In unsupervised learning, the model extracts relationships in data. As opposed to supervised learning, unsupervised learning uses only input data without any labeled outputs. While unsupervised learning has many forms, practitioners often encounter two primary challenges: clustering and density estimation. Clustering involves finding groups in data while density estimation involves summarizing the distribution of data. There are additional unsupervised methods that can be employed, such as visualization methods, which involve presenting data in different ways, and projection methods, which reduce the dimensionality of the data. A reinforcement learning problem is one in which a learning agent operates in a complex environment while receiving feedback in order to improve its performance. The goal of reinforcement learning is to maximize a numerical reward signal by mapping situations to actions. A learner is not told which actions to take but must discover which actions yield the best rewards through trial and error (Biship and M. 2017; Goodfellow et al. 2016; Norvig and Russell 2020).

Deep learning is a type of machine learning algorithm with a brain-like logical structure of algorithms called artificial neural networks. Some of the key differences between machine learning and deep learning are as follows: First and foremost, while traditional Machine Learning algorithms have a rather simple structure, such as linear regression or a decision tree, deep learning is based on an artificial neural network. Secondly, deep learning algorithms with self-learning capabilities require much less human intervention and bypass manual feature engineering steps required in machine learning. Finally, deep learning requires more data than a traditional Machine Learning algorithm to function properly and unlike machine learning, its performance will improve as it receives more data.

In biomechanics and gait analysis, machine learning has depicted numerous applications in recent years. These applications include but are not limited to diagnosis of gait disorder, activity classification and recognition, predicting early intervention for fall-related risks due to a disability or aging (Begg et al. 2005; Paulo et al. 2019; Yoo et al. 2013; Zhou et al. 2020b), determining motor recovery tasks (Goh et al. 2018), or planning rehabilitation or therapeutic interventions (Liu et al. 2016). The most common area of application has been the classification of movement patterns, with many studies focusing on distinguishing pathological kinematics from normal kinematics (Halilaj et al. 2018). In gait analysis, supervised learning is increasingly used to model biomechanical systems $Y=G(x)$ by determining their relationship between inputs X and outputs Y. It usually involves raw multidimensional arrays [Ti, Fi], where Ti represents a number of subjects or their trails, and Fi represents data features such as wearable sensor signals, kinematics,

kinetics, or neuromuscular signals. The model output can be either groups or categories, such as gait events, activities, or disorders, or it could be representative of numerical values such as gait parameters, joint kinematics, force, or moments (Khera and Kumar 2020).

2.3 Spatial Temporal Prediction

There are many factors that can be used to quantify an individual's gait, including spatial and temporal parameters. Spatial parameters refer to the distance between the feet while walking, while temporal parameters refer to the timing of each step. These factors can be affected by a variety of things, such as age, weight, and terrain. For example, older adults tend to have shorter strides and take more time between steps than younger adults. Heavier individuals tend to have shorter strides and take more time between steps than lighter individuals. Walking on uneven or slippery surfaces can impact one's gait by causing them to slow down or change their stride pattern. There is a growing body of evidence that suggests spatial-temporal parameters derived from gait analysis and IMUs can provide valuable insights into an individual's movement patterns. These simple metrics can offer clinical value for diagnosis and monitoring disease progression. Several studies have attempted to validate IMUs in both the gold-standard system along with the current clinical methods (Hamidon et al. 2022; Piche et al. 2022; Washabaugh et al. 2017). These studies can be categorized into two groups, deterministic and probabilistic (or stochastic). In the determinist groups, researchers rely heavily on the previously mentioned assumptions (e.g. biomechanical models, double integration, etc) to calculate spatial-temporal parameters. On the contrary, probabilistic studies use a data-driven approach to develop the model and estimate these parameters.

Several studies calculate spatial-temporal gait parameters by reconstruction of foot trajectories through gait phase identification and then double integration of the linear accelerations measured by IMUs. Sensor fusion techniques (Bertoli et al. 2018; Kluge et al. 2017; Rampp et al. 2014; Trojaniello et al. 2014) and extended Kalman filters (Bailey and Harle 2014; Foxlin 2005; Zizzo and Ren 2017) are commonly used to reduce noise, improve the accuracy in gait phase identification, and compensate for drift (Bertoli et al. 2018). For instance, Ramp et al, used two-foot IMUs to measure spatial-temporal gait parameters for elderly subjects with and without a wheeled walker. First, they used dynamic time wrapping to segment the gait cycle, then identified gait events such as heel-strike and toe-off via zero crossing of the gyroscope's medial-lateral acceleration signal to calculate temporal parameters. To calculate stride length, the gravity-compensated accelerometer signal was double integrated, and sensor drift was modeled using a piece-wise defined linear function. Stride length and stride time showed a correlation of 0.93 and 0.95 with reference measurements, and the absolute error of stride length was 6.26 cm during normal walking (Rampp et al. 2014). In a similar study, Trojaiello, et al. used the cyclic nature of gait to identify gait events and subsequently calculated the spatial and temporal parameters. They reached an average accuracy and precision of 0.1±1.9 cm and achieved their desired precision of ±1 cm (Trojaniello et al. 2014). Ferrari et al. reported measurement errors of −0.16±7.02 cm for stride length based on double integration (Ferrari et al. 2016).

For data-driven methods, such as machine learning or deep learning, many studies attempted to estimate spatial-temporal gait parameters across various activities, patient

populations, and sensor configurations. Aminian et al. in 1995 used a regression model to link the raw sensor data of gait directly with the corresponding spatial-temporal parameters. They used a two layer perceptron model to estimate speed during walking (Aminian et al. 1994). With the recent advances in deep learning, parametrization of input data has become obsolete. Deep learning techniques allow for more accurate predictions and classification of data. User-friendly interfaces make it easy to implement these techniques via libraries such as Pytorch or TensorFlow without the need for extensive knowledge of programming. Moreover, a significant amount of data has become available due to advancements in sensors and cloud technology. As a result, deep learning is becoming increasingly popular to analyze this data for tasks such as image recognition, natural language processing, and time series forecasting. The true potential of neural networks can be exploited by analyzing raw sensor data rather than parameters calculated from the sensor data and by employing sufficiently deep architectures. In 2016, Hannink et al used a convolutional neural network model for stride length estimation from raw IMU data. They used a publicly available and clinically relevant benchmark dataset consisting of 1220 strides from 101 geriatric patients. Their model yielded performance with a mean accuracy and precision of 0.01±5.37 cm (Hannink et al. 2016a). Their results were considerably more accurate than previous integration-based methods (Rampp et al. 2014). Moreover, stride length predictions were robust to different methods of stride segmentation, improving the clinical applicability for patients with pathologic gait (Hannink et al. 2018). In another study, the same authors extracted eight spatial-temporal gait parameters using a similar approach and reported stride length, width and internal-external foot rotation with accuracies of -0.15 ± 6.09 cm,

−0.09 ± 4.22 cm and 0.13 ± 3.78° respectively. Stride, swing and stance time, along with heel and toe contact times, were estimated with accuracies of ±0.07, ±0.05, ±0.07, ±0.07 and ±0.12 s respectively (Hannink et al. 2016b).

2.4 Joint Kinematic Prediction

The first study estimating joint kinematics using inertial sensors dates back to 1990 (Willemsen et al. 1990). Since then, various methodological approaches have been presented to estimate 2D and 3D joint kinematics using wearable inertial sensors. Regarding the estimation of 2D joint kinematics, Picerno et al. described four approaches, three deterministic and one data-driven approach. Deterministic methods included estimating the joint kinematic by comparing the equivalent accelerations of the proximal and distal body segments at the connecting hinge joint (Dejnabadi et al. 2005; Willemsen et al. 1990), by comparing the planar orientations of two adjacent body segments, and by combining the two methods (Findlow et al. 2008; Seel et al. 2014). Similar to spatial-temporal parameter estimation, the data driven methods were developed by machine learning or neural networks (Findlow et al. 2008). In particular, Seel et al. (Findlow et al. 2008; Seel et al. 2014) proposed a method that used gyroscopes and accelerometers to first define the knee joint axis and position, and then calculate the flexion/extension joint angle using a Kalman filter. They reported errors of less than 1° against a standard motion analysis system. Due to the fact that sensor-to-segment axis alignment was performed using ad hoc segment rotational movements with the direction of the angular velocity vector as the joint rotation axis, the IMU could be positioned anywhere on the body segment. Despite their innovation in estimating the position of the joint rotation center with respect to the

17

sensor from arbitrary movements, they assumed the knee joint behaved like a mechanical hinge joint. In reality, the knee joint has other degrees of freedom such as internal/external and varus/valgus rotations, as well as small translational movements.

For 3D joint kinematic measurements, sensor fusion algorithms of IMU signals including magnetometer data have been used to estimate the sensor's 3D orientation in space. Calibration methods are used to estimate orientations of bone coordinate systems with respect to the sensor attached to the body segments (Cutti et al. 2009; Favre et al. 2009; O'Donovan et al. 2007). Picerno et al calculated the 3D joint kinematics of the hip, knee, and ankle during gait. They reported RMSE of 1.9°, 2.8°, and 3.6° on lateral, frontal, and transverse plane rotations respectively (Picerno et al. 2008). More recently, the accuracy of IMU-based kinematic measurements has been improved by integration with an optimization algorithm to impose realistic joint constraints to the estimated movement (Karatsidis et al. 2018). The sagittal plane joint angles of ankle, knee, and hip presented excellent Pearson correlations ($\rho = 0.95$, 0.99, and 0.99, respectively) and RMSE of $4.1 \pm 1.3°$, $4.4 \pm 2.0°$, and $5.7 \pm 2.1°$, respectively. The hip internal rotation indicated the least accurate degree of freedom with RMSE $6.5 \pm 2.8°$ and Pearson correlations of 0.68.

These methodologies, however, require nontrivial computational resources, are not robust to sensor placement, require multiple IMUs and complex coordinate transformation, and are limited to only planar joint angles, making them less suitable for real-time applications with instantaneous feedback. The first attempt for using data-driven methods to estimate joint kinematics was introduced by Findlow and colleagues. In their study, they used a regression algorithm to train a neural network for predicting transverse plane joint

kinematics of the lower limb from IMUs. They achieved the best result for intra-subject predictions (MSE=2.3°, r=0.99) and the worst results for inter-subject predictions (MSE=7.8°, r=0.88) (Findlow et al. 2008). More recent studies have tried advanced machine learning methods such as deep neural networks to estimate joint kinematics using single and multiple sensors (Argent et al. 2019; Hernandez et al. 2021; Hossain et al. 2022; Lim et al. 2019; Mundt et al. 2019; Tan et al. 2022; Weygers et al. 2020). For instance, Gholami et al. trained a convolutional neural network based on a single IMU and estimated the joint angles of the hip, knee, and ankle with RMSE of less than 3.5° and 6.5° in intra- and inter-subjects evaluations (Gholami et al. 2020). Hernandez et al 2019, used the CNNLSTM model and five IMU sensors located on the pelvis, thighs, and shanks to estimate the lower limb joint kinematics across walking, running, and transition. They reported MAE for the DOFs ranged from 2.2(0.9)° to 5.1(2.7)° with an average of 3.6(2.1)° (Dorschky et al. 2020; Hernandez et al. 2021). In terms of 3D knee joint prediction, only a handful of studies have attempted to use the same techniques to estimate knee varus/valgus or internal/external rotations (Mundt et al. 2019, 2020b; Stetter et al. 2020). Mundt et al used simulated inertial sensor data, including linear accelerations and angular rates, as an input for training deep neural networks to predict joint angles and moments of the lower limbs during gait. In minor motion planes, they achieved correlation coefficients exceeding 0.80 and in sagittal plane, they achieved correlation coefficients exceeding 0.98 (Mundt et al. 2019).

In a clinical study, the feasibility of using deep learning and IMUs for subjects with knee osteoarthritis performing multiple clinically important activities to predict knee joint

sagittal plane kinematics was assessed (Tan et al. 2022). They trained a bidirectional long short-term memory model on IMU data and estimated knee joint flexion kinematics for phases of walking, transitioning into and out of a chair, and negotiating stairs. Across the different activities, RMSE (SD) ranged from 7.04° (2.6) to 11.78° (6.04), and Pearson's R from 0.85 to 0.99.

Although these studies indicated great potential for using deep learning to assess joint kinematics from IMUs, the accuracy of these algorithms relies on large and representative biomechanics training datasets that are frequently expensive and time-consuming to collect. To expand the availability of training data, researchers are leveraging artificially generated data to improve model prediction accuracy and reliability. The most common technique to generate artificial IMU data in movement analysis is to leverage existing passive-marker motion capture datasets to calculate simulated IMU data based on marker trajectories and accelerations (Dorschky et al. 2020; Mundt et al. 2019, 2020b; Young et al. 2014). Mundt et al 2020 reported a mean correlation coefficient of 0.85 for the joint angles and a maximum RMSE of 4.3° and showed a slight improvement in joint kinematic prediction with increasing correlation coefficient by 0.04 and decreasing RMSE by 0.5° (Dorschky et al. 2020; Mundt et al. 2019, 2020b; Young et al. 2014). In another study, Dorschkey et al also supported this claim and reported a decrease in the root mean square error (RMSE) of the hip, knee, and ankle joint angles up to 17%, 27%, and 23% when the deep learning model is trained on both simulated and measured data (Dorschky et al. 2020; Mundt et al. 2019, 2020b; Young et al. 2014).

Overall, data-driven methods such as machine learning indicated promising results in human movement analysis in recent years. They can be leveraged to bypass the need for complex coordinate transformations and sensor-to-segment alignments, reducing error, computation complexity, and time, making them an ideal method for future research. Machine learning algorithms can be used in biomechanics to estimate nonlinear relationships between inputs and outputs, like those existing between segment IMU sensors data as inputs and kinematics and GRFs, joint moments, and patient outcome as outputs.

2.5 Gaps and Opportunities

By 2040, 78 million Americans will suffer from osteoarthritis. This is a leading cause of disability with a projected cost of $128 billion per year (Barbour et al. 2017; Hootman et al. 2016). Total knee replacements are a treatment for chronic knee pain secondary to osteoarthritis. According to a 2014 study, one-third of knee replacement patients may not have been appropriate candidates for the procedure due to arthritis symptoms not being severe enough to merit the need for aggressive intervention. With a typical 20-year lifetime, there is an advantage to the patient in delaying the initial procedure to avoid a second revision surgery to replace the worn knee replacement. Access to enhanced diagnostic information using simple in-home motion tracking would improve the monitoring of disease progression along with rehabilitation for many movement disorders, including osteoarthritis and total joint replacement (Ramkumar et al. 2019). The technology development for wearable devices is one such automated technology to improve health care, address health care workforce shortages, reduce costs, and enable personalized medicine tailored to individuals. In this regard, a smart patient movement monitoring

system using wearables (IMUs) and machine learning offers an untapped area of potential research growth, but progress needs to be made to increase usability(Appelboom et al. 2014; Kurtz et al. 2022; Nahavandi et al. 2021). This opportunity can be broken down into multiple distinct areas such as dataset development, activity classification, movement parameters estimation, synthetic data generation, universal model development, and clinical application.

2.5.1 Datasets

Training and development of ML algorithms require large datasets. This dataset should include diverse patient populations, various clinical or routine activities of daily living, wearable sensor data (e.g. IMUs), RGB video, and desired biomechanics output such as movement parameters or patient outcomes. Due to the cumulative nature of research, open datasets can have a significant impact on advancing the state of the art. Scientific advances are greatly accelerated by these resources because they foster new analyses, improve data practices, and facilitate reproducibility. The oldest biomechanics-related datasets go back to 1983 when David A. Winter released a 2-D walking dataset (Winter 1983). Advances in data-driven modeling has outpaced the available data such that new and open resources are critical to continued progress. Some recent efforst to address this challenge includes open datasets of EMG, IMU, goniometer, force plate, and MOCAP data for various locomotion activities from Moore et al. (2015), R.K. Fukuchi et al. (2017), Hu et al. (2018), C. A. Fukuchi et al. (2018), Schreiber and Moissenet (2019), and Lencioni et al. (2019). A more recent dataset was presented by Camargo et al 2021, which offers a comprehensive source of locomotion information. This dataset contains 3-dimensional biomechanical and

wearable sensor data from 22 able-bodied adults for multiple locomotion modes (e.g., level-ground/treadmill walking, stair ascent/descent, and ramp ascent/descent) and multiple terrain conditions of each mode (walking speed, stair height, and ramp inclination). In addition to the sensor output, the dataset includes kinematics and kinetics of the lower limbs. However, to our knowledge, there has not been any equivalent datasets made available for patients with knee OA or total knee replacement.

2.5.2 Movement Parameters Estimation

Movement parameters can be categorized into spatial-temporal parameters, joint kinematics, or joint moments. Spatial-temporal gait parameters can provide valuable clinical insight regarding gait patterns, risk of falling (Weiss et al. 2013), and disease progression (Hausdorff 2005) in osteoarthritis patients and recovery for patients with a total knee replacement (Levinger et al. 2013; Outerleys et al. 2021; Wu et al. 2017). Previous studies indicated that incorporating IMUs on additional body segments may provide access to additional gait metrics and improve gait characteristic predictions, especially for patient populations with pathologic movement characteristics (Atallah et al. 2012; Carcreff et al. 2018; Zijlstra and Hof 2003). For instance, Carcreff et al. demonstrated that IMUs placed on the shank and thigh yielded more accurate predictions of stride time, length, and velocity compared to feet-mounted IMUs for children with cerebral palsy, particularly for those patients with increased disability (Atallah et al. 2012; Carcreff et al. 2018; Zijlstra and Hof 2003). Patients with progressive OA typically exhibit gait adaptations including decreased joint flexibility, increased stance time on the affected side, cadence, and double support time, and an overall increase in variability of these spatial-temporal parameters (Bejek et

al. 2005; Hollman et al. 2011; Kiss 2011; Kiss et al. 2012; Zeni and Higginson 2009). It is unclear how these gait adaptations progress over time and impact the prediction of gait mechanics using IMUs. Additionally, systematic studies quantifying optimal sensor combinations for the best performance across various patient populations are important to this field but are lacking. Specifically, the feasibility of estimating spatial-temporal gait parameters for patients with OA and TKA and the optimum sensor configuration for this patient cohort has not been investigated.

Several efforts have been made to estimate joint kinematics by leveraging deep learning methods. The majority of these studies used simulated IMUs to estimate joint kinematics from existing motion capture data, were limited to only sagittal plane joint angles during normal gait, and the data was collected only from healthy populations. Unlike deterministic methods, the accuracy of data-driven algorithms relies on large and representative biomechanics training datasets (Argent et al. 2019; Gholami et al. 2020; Halilaj et al. 2018; Lim et al. 2019; Wouda et al. 2018). Models trained on healthy populations may not necessarily be functional for pathological movement patterns in OA or TKA patients. To our knowledge, there has not been any study that uses machine learning methods to estimate joint kinematics from IMUs for patients with OA and TKA across various activities.

2.5.3 Synthetic Data Generation

Deep learning-based methods have gained popularity in recent years in many fields, including biomechanics, but are data-hungry algorithms. To develop a reliable machine learning/deep learning model, it's necessary to have sufficient clean and representative

data. However the process of data collection in the biomechanics field is extremely expensive, time-consuming, and difficult when patient privacy is involved (Murdoch 2021; Tobore et al. 2019). To expand the availability of training data, researchers are leveraging artificially generated (synthetic) data to improve model prediction accuracy and reliability (Chen et al. 2021). One of the most common techniques to generate artificial IMU data in movement analysis is to leverage existing passive-marker motion capture datasets to calculate simulated IMU signals from marker trajectories (Brunner et al. 2015; Johnson et al. 2020; Young et al. 2014). Using this technique, Mundt et al. (Mundt et al. 2020b) combined simulated IMU data from an archived MOCAP database with experimentally measured IMU data on a smaller subject cohort to predict lower limb kinematics and kinetics during gait. They showed the inclusion of the simulated data in the training set reduced the root mean square error in joint kinematic estimates from 4.8° to 4.3° but did not improve joint kinetics predictions. One limitation of this method is that the simulated IMU signals are limited to only existing movement data. In another study, Dorschky et al (Dorschky et al. 2020) combined measured IMU data from subjects during walking and running with artificial IMU data generated from complementary musculoskeletal models (MSMs). The authors applied perturbations to the MSM's joint angles, ground reaction forces, and speeds based on random sampling from the experimental measures to generate synthetic IMU data for movements not observed experimentally. The constraints of the MSM and corresponding optimal control algorithm ensured the perturbations resulted in physically realistic joint mechanics. Using this technique, they were able to generate artificial IMU data for movements beyond those observed experimentally. Similar to

25

Mundt et al. (Mundt et al. 2020b), the addition of the synthetic data improved kinematic predictions at the hip, knee, and ankle. Both techniques rely on intensive gait lab data collection, limiting widespread accessibility. In contrast, synthetic data could be generated with only a few representative gait lab measurements of an activity to establish the general kinematic patterns of the movement. It has yet to be demonstrated if machine learning models can achieve the necessary accuracy when trained exclusively with synthetic data.

In this landscape, there is a branch of deep learning which is a hybrid of supervised and unsupervised learning methods. The goal of this research is to develop a deep learning-based method for generating new synthetic data. Generative Adversarial Network (GAN) (Goodfellow et al. 2014), Variational Autoencoder (VAE) (Kingma and Welling 2013), flow-based models (Rezende and Mohamed 2015), and diffusion-based models (Sohl-Dickstein et al. 2015) are some of the popular models when it comes to generating images and sequences. Under various autoencoder groups, there's a linear autoencoder model which is simple, fast, and reliable and can be used for reversible mapping of input data to latent space. This mapping could be conventionally done through linear matrix operations, such as principal component analysis (PCA). PCA analysis not only can reveal information about data via dimension reduction but it can also be used as an autoencoder model to generate synthetic data. For instance, Jung-Hoon Kim et al (Kim et al. 2021) used this method for reconstructing fMRI data. However, the application of PCA to generate patient-specific OA and TKA kinematics data across various activities has not been demonstrated.

Therefore, in the next chapters, we aim to tackle these challenges and fill in the gaps in existing researches, starting with creating comprehensive datasets for OA and TKA

26

patients, developing effective methodologies for creating an AI-driven tools, and

proposing universal model for monitoring patient movement in the future studies.

CHAPTER 3: DEEP LEARNING IN GAIT PARAMETER PREDICTION FOR OA AND TKA PATIENTS WEARING IMU

3.1 Abstract

Quantitative assessments of patient movement quality in osteoarthritis (OA), specifically spatiotemporal gait parameters (STGPs), can provide in-depth insight into gait patterns, activity types, and changes in mobility after total knee arthroplasty (TKA). A study was conducted to benchmark the ability of multiple deep neural network (DNN) architectures to predict 12 STGPs from inertial measurement unit (IMU) data and to identify an optimal sensor combination, which has yet to be studied for OA and TKA subjects. DNNs were trained using movement data from 29 subjects, walking at slow, normal, and fast paces and evaluated with cross-fold validation over the subjects. Optimal sensor locations were determined by comparing prediction accuracy with 15 IMU configurations (pelvis, thigh, shank, and feet). Percent error across the 12 STGPs ranged from 2.1% (stride time) to 73.7% (toe-out angle) and overall was more accurate in temporal parameters than spatial parameters. The most and least accurate sensor combinations were feet-thighs and singular pelvis, respectively. DNNs showed promising results in predicting STGPs for OA and TKA subjects based on signals from IMU sensors and overcomes the dependency on sensor locations that can hinder the design of patient monitoring systems for clinical application.

3.2 Introduction

Quantitative assessments of movement quality in osteoarthritic (OA) and joint reconstruction patients, specifically spatial-temporal gait parameters (STGPs), provide valuable insight into gait patterns, activity type (Witjes et al. 2016), risk of falling, and disease progression (Hausdorff 2005; Weiss et al. 2013) . This diagnostic information is used in a number of applications that include development of personalized treatment plans, optimized post-operative rehabilitation, monitoring changes in mobility of patients after surgery (Fransen et al. 2017; Lee et al. 2015; Levinger et al. 2011; Snell et al. 2018), advancement of promising new interventions, and reducing overall medical costs (Hausdorff 2005; Weiss et al. 2013). Conventional methods for measuring gait characteristics that include motion capture (MOCAP) systems and force plates require a laboratory environment and expensive, time-consuming, equipment (Hannink et al. 2017). On the contrary, wearable sensors, specifically inertial measurement units (IMUs), are lightweight, inexpensive, and mobile. IMU's measurement fidelity has improved significantly in recent years and have been used in various applications including 3D character animation, robotics, automotive vehicles, drones, and human motion measurement (Imtiaz et al. 2014) .

Processing streams of IMU data to extract clinically meaningful movement characteristics, such as activity classification, spatial-temporal parameters, gait pathology, and gait phase detection is challenging (Nweke et al. 2018). Several studies calculate spatial-temporal gait parameters by reconstruction of foot trajectories through double integration of the linear accelerations measured by IMUs. Sensor fusion techniques (Bertoli

29

et al. 2018; Kluge et al. 2017; Rampp et al. 2014; Trojaniello et al. 2014) and extended Kalman filters (Bailey and Harle 2014; Foxlin 2005; Zizzo and Ren 2017) are commonly used to reduce noise and improve measurement accuracy. These methods rely on identification of the zero-velocity condition of the foot during stance for gait segmentation. However, clear zero-velocity conditions are difficult to identify for patients with pathological gait or during highly dynamic activities like free running (Zrenner et al. 2018).

Data-driven approaches like deep learning have shown promising results in extracting complex patterns from data in the fields of computer vision, speech-recognition, and sequence modeling. Researchers have used deep learning on IMU-based movement data to classify different activities or quantify activity-specific movements (Hu et al. 2018; Kautz et al. 2017; Ordóñez and Roggen 2016; Zheng et al. 2018). Hannink et al. demonstrated the ability of deep-learning algorithms to recognize the non-linear relationships between raw IMU data and stride length as well as other STGPs (Hannink et al. 2016b) . Using a deep convolutional neural network trained on over 1220 strides from 101 geriatric patients, the algorithm predicted stride length with a mean error of −0.15 cm, which was considerably more accurate than previous integration-based methods(Rampp et al. 2014) . Moreover, stride length predictions were robust to different methods of stride segmentation, improving the clinical applicability for patients with pathologic gait (Hannink et al. 2018). Similar results have been demonstrated using neural networks for measuring stride length during free running(Zizzo and Ren 2017) , but variability in foot strike patterns (e.g., heel strike versus toe strike) reduced accuracy highlighting the importance of population-specific datasets for best results.

30

Single-body segment mounted IMUs (e.g., wrist or pelvis) are limited in calculation of certain STGPs such as number of steps, step cadence, or step distance which may not be adequate for clinical applications (Fasel et al. 2017; Soltani et al. 2018). Incorporating IMUs on additional body segments (e.g., foot, shank, thigh, pelvis, or trunk) may provide access to additional gait metrics and improve gait characteristic predictions, especially for patient populations with pathologic movement characteristics (Atallah et al. 2012; Carcreff et al. 2018; Fasel et al. 2017; Zijlstra and Hof 2003). Carcreff et al. demonstrated that IMUs placed on the shank and thigh yielded more accurate predictions of stride time, length, and velocity compared to feet mounted IMUs for children with cerebral palsy, particularly for those patients with increased disability(Carcreff et al. 2018). Patients with progressive OA typically exhibit gait adaptations including decreased joint flexibility, increased stance time on the affected side, cadence, and double support time, and an overall increase in variability of spatial temporal parameters (Bejek et al. 2005; Hollman et al. 2011; Kiss 2011; Kiss et al. 2012). It is unclear how these gait adaptations progress over time and impact the prediction of gait mechanics using inertial sensors. Additionally, systematic studies quantifying optimal sensor combinations for the best performance across various patient populations are important to this field, but are lacking.

Thus, the purpose of this study was two-fold: (1) to access the ability of multiple contemporary deep neural network architectures to predict STGPs from IMU data in the OA and joint-replacement patient populations and (2) to determine the optimal sensor combination to maximize prediction accuracy. The results of this study will help patients suffering from OA who may go on to receive a total joint replacement benefit from the

accurate real-time patient monitoring of STGPs to inform their treatment, surgical planning, and rehabilitation.

3.3 Methods

3.3.1 Gait Measurements of Osteoarthritic and Total Knee-Replacement Subjects

Twenty-nine subjects, including 14 subjects with OA (Age = 67 ± 7, weight = 79 ± 12 kg, height = 168 ± 16 cm, 4 females and 10 males), 15 subjects with total knee arthroplasty (TKA) (Age = 68 ± 4, weight = 76 ± 14 kg, height = 164 ± 9 cm, 11 females and 4 males, 7 uni-lateral and 8 bi-lateral), participated in the study as part of a larger investigation. All participants signed a consent form prior to the experiment with IRB approval (# 1328728). Subjects were fitted with 71 reflective markers on anatomical landmarks and 17 IMUs on various limb segments and the trunk. For this study, only the 7 IMUs located on the feet, shanks, thighs (Patterson et al. 2016; Vargas-Valencia et al. 2016), and pelvis (Bolink et al. 2012) were used in the subsequent data analysis (Figure 1a,b). Subjects performed 15 trials of a 5-m walking task at three different speeds: self-selected, slow, and fast to cover the entire range of possible daily walking paces. During fast walking, subjects were instructed to walk at their maximum comfortable speed without running (brisk walking) typified by longer steps at a faster cadence. During slow walking, subjects were instructed to walk at their slowest speed, typified by shorter steps at a slower cadence. During the walking tests, synchronized data was collected from a 13 camera Vicon motion capture system (Centennial, CO), 4 Bertec force platforms (Columbus, OH), and IMUs (Xsens, Enschede, Netherlands) (Figure 1a). The sampling frequency of force data, MOCAP, and

IMUs (free acceleration and angular velocity) were 1000 Hz, 100 Hz, and 40 Hz, respectively.

3.3.2 Gait Data Processing

MOCAP data were segmented into a full stride for each leg based on two successive heel strikes identified using the heel markers' vertical position (Fellin et al. 2010). For each full stride, the heel strike and toe off times, spatial characteristics (step length, stride length, step width, and toe out angle), temporal characteristics (step time, stride time, stance time, swing time, single support time, and double support time), and general characteristics (cadence and speed) were calculated (Panero et al. 2018; Veilleux et al. 2016).

IMU data for each trial was up-sampled to 100 Hz and segmented into full strides for each leg based on the angular velocities of the feet sensors in the sagittal plane using the peak detection method (Figure 3.1c,d) (Barth et al. 2015; Ghassemi et al. 2018). The mean absolute error between heel strike and toe-off events identified using the IMU and MOCAP data were 0.02 ± 0.01 and 0.04 ± 0.01 s, respectively. Linear accelerations and angular velocities from the IMU-based coordinate systems for left legs were reflected about the medio-lateral axis to provide consistent anatomical directions for left and right limb segments. The IMUs' six channels of acceleration and angular velocity data were normalized using the maximum sensor acceleration and angular velocity range. A zero-padding technique was used to ensure the IMUs' data sets had a consistent length of 212 points prior to use in the deep-learning models (Hannink et al. 2016b). The IMU data for each stride segment was labeled with the gait characteristics calculated using the MOCAP data for use in the subsequent supervised machine learning models. The gait data

processing yielded 3778 segmented and labeled strides from the 29 subjects. A descriptive

statistical analysis was conducted on measured spatial, temporal, and general gait

parameters to characterize the dataset. This includes mean, standard deviation, coefficient

of variation, and interquartile range for knee OA and TKA subject cohorts at three paces,

slow, normal, and fast.
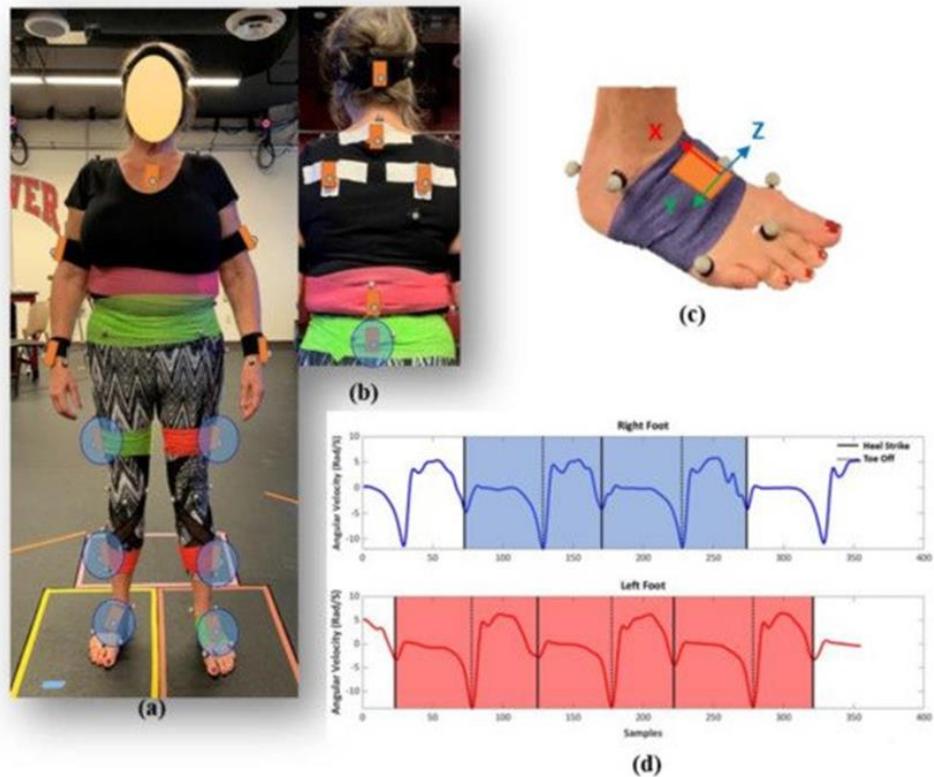
Figure 3.1: Subject suited up with markers and inertial measurement units (IMUs) (a) front, and (b) back view. IMUs circled in blue (feet, shanks, thighs, pelvis) were used in the supervised machine learning models. (c) IMU sensor attached on right foot with coordinate system, (d) a sample of segmented IMU data based on angular velocities of feet sensors.

3.3.3 Preliminary Neural Network Architecture Benchmarking and Selection

Six contemporary multivariate time series neural network architectures were utilized to predict stride length from our subject cohort based solely on the feet IMU data (Table 3.1). Stride length and the feet IMUs were chosen to enable benchmarking prediction accuracy against published studies. For network training, 80% of strides from 26 of 29 subjects were randomly allocated to the training set and the remaining 20% of strides from the same subjects were allocated to a validation set. Strides from the final three subjects not included in the training set, one OA subject, one uni-lateral TKA subject, and one bilateral TKA subject, were allocated to a test set. Network prediction accuracy was assessed using 5-fold cross-validation, with training, validation, and test sets randomly reallocated for each fold of the cross-validation. Optimal architecture with the lowest errors for both validation and test sets was selected for conducting a design of experiment on prediction of STGPs with different sensor numbers and locations.

Table 3.1: Contemporary multivariate time-series deep-learning models for prediction of stride length.

Reference	Models
Hannink 2017 (Hannink et al. 2016b)	Convolutional Neural Network (CNN)
Zrenner 2018 (Zrenner et al. 2018)	Convolutional Neural Network (CNN)
Wang 2017 (Wang et al. 2017)	Fully Convolutional Networks (FCN)
Wang 2017 (Wang et al. 2017)	Residual Network (ResNet)
Karim 2019 (Karim et al. 2018)	Multivariate Long Short-Term Memory Fully Convolutional Network (MLSTM-FCN)

Neural networks were trained using a backpropagation and stochastic gradient descent optimization approach to minimize the loss function, mean square error (MSE), between the model-predicted and labeled stride length, using the form:

where $\hat{y}_i$ was the model predicted stride length, y_i was the labeled stride length, and n was the total number of strides in the training set.

$$MSE = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n} \qquad (1)$$

An adaptive learning rate optimization with a learning rate, beta-1, and beta-2 of 0.001, 0.9, and 0.999, respectively were used for training all networks (Kingma and Ba 2014) with a total epoch of 300. Once each network was trained, the predictive accuracy was quantified by calculating the mean error (ME) and the mean absolute error (MAE) between the predicted and measured stride lengths for both the validation and test sets. ME was calculated to enable comparison with previously published studies, whereas the MAE provides a better metric for true prediction accuracy. To enable an equitable comparison of the prediction accuracy across various gait characteristics with different magnitudes and units, the absolute error was divided by the mean of the labeled test data resulting in the normalized absolute percent error (NAPE).

3.3.4 Assessing Optimal Sensor Combinations for Each Gait Characteristic

Based on the result of the preliminary neural network architecture selection, the 1D convolution neural network (CNN) architecture proposed by Zrenner et al. was chosen for a larger design-of-experiment study on sensor combinations (Zrenner et al. 2018) . This network consisted of two convolutional layers followed by two max pooling layers, a flattening layer, and two fully-connected layers. Rectified linear unit (ReLu) activation functions were placed after each layer. Keras with a Tensorflow backend was used for training the architecture (Abadi et al. 2016; Chollet and others 2015).

A full factorial design of experiments was implemented to analyze the prediction accuracy based on 15 unique combinations of the feet, pelvis, shank, and thigh sensors (Table 3.2). Leveraging the ensemble approach proposed by Hannick et al., individual CNNs were trained using the segmented and labeled stride IMU data to predict each of the 12 spatial, temporal, and general gait parameters (Figure 3.2) for each unique sensor combination(Hannink et al. 2016b). The same training set definitions, 5-fold cross-validation, and training approaches were used as in the preliminary analysis. Likewise, the same MAE and NAPE error estimations were calculated for each gait parameter with each sensor combination.

The Friedman test, which is a non-parametric statistical test analog to a repeated measures analysis of variance (ANOVA), was used to detect statistically significant differences in prediction accuracy (NAPE) across sensor combinations. Stepwise Dunn's post hoc tests followed by Bonferroni correction due to multiple testing was performed to establish significant differences (new p-value: $0.05/105 = 0.000476$). To determine an

overall optimal sensor combination, sensor combinations were ranked based on Friedman

ranking and averaged across all the gait parameters for each sensor combination(Trawiński

et al. 2012).

Table 3.2: Sensor combinations used in the design of experiment.

n	Feet	Pelvis	Shank	Thigh	Combinations
1	×				Feet (F)
2		×			Pelvis (P)
3	×	×			Feet Pelvis (F P)
4			×		Shank (S)
5	×		×		Feet Shank (F S)
6		×	×		Pelvis Shank (P S)
7	×	×	×		Feet Pelvis Shank (F P S)
8				×	Thigh (T)
9	×			×	Feet Thigh (F T)
10		×		×	Pelvis Thigh (P T)
11	×	×		×	Feet Pelvis Thigh (F P T)
12			×	×	Shank Thigh (S T)
13	×		×	×	Feet Shank Thigh (F S T)
14		×	×	×	Pelvis Shank Thigh (P S T)
15	×	×	×	×	Feet Pelvis Shank Thigh (F P S T)
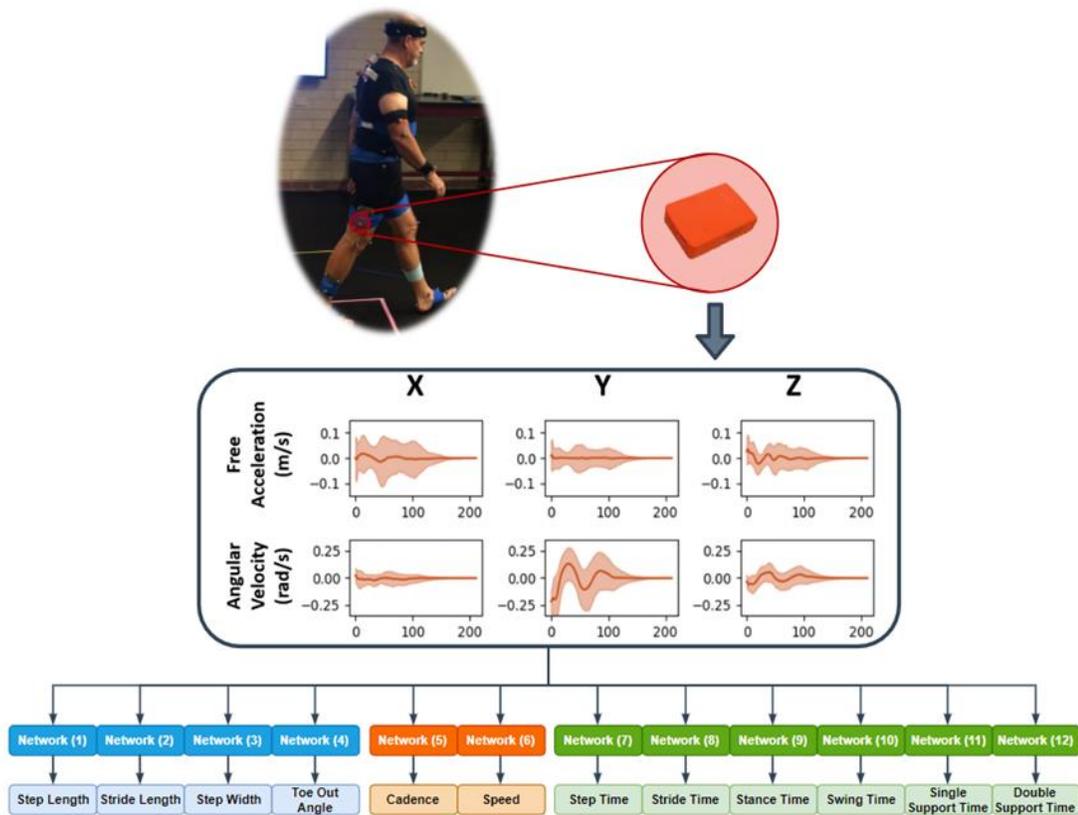
Figure 3.2: Workflow from in-vivo to spatial temporal gait parameter prediction.

3.4 Results

3.4.1 Spatial-Temporal Gait Parameters (STGPs) Statistical Analysis

The OA group demonstrated larger step width (+2.8 cm) and toe out angle (+4.9 deg), as well as smaller step length (−1.9 cm), stride length (−3.9 cm), double support time (−0.1 s), and speed (−3.4 cm/s) on average for three different paces compared to the TKA group.

In general, OA patients demonstrated greater variation (standard deviation (SD), coefficient of variation (CV), and range) in all but two of the STGPs measured compared to TKA patients. Increases in variability (SD) was also observed for step length, stride

length, cadence, and speed for fast trials in both OA and TKA groups compared to

normal and slow paces (Table 3.3).

Table 3.3: Descriptive statistic of spatial, temporal, and general parameters of dataset grouped by knee and pace (SD: Standard Deviation, CV: Coefficient of Variation, IQR: Interquartile Range).

Variable	Knee Status	Pace	Mean	SD	CV	Range	IQR
Step Length (cm)	OA	Fast	66.6	10.4	15.6	50.0	9.0
		Normal	56.7	9.1	16.1	54.6	9.2
		Slow	53.2	6.9	13.0	45.0	9.8
	TKA	Fast	66.0	9.9	15.1	46.7	10.8
		Normal	59.1	7.6	12.9	52.3	10.7
		Slow	53.7	6.5	12.1	37.1	8.4
Stride Length (cm)	OA	Fast	132.9	20.3	15.3	91.5	16.3
		Normal	113.0	17.4	15.4	88.4	17.0
		Slow	106.1	12.9	12.1	67.7	18.2
	TKA	Fast	132.1	18.9	14.3	93.4	17.4
		Normal	118.1	14.3	12.1	71.4	19.8
		Slow	107.2	12.3	11.5	67.2	14.5
Step Width (cm)	OA	Fast	13.2	5.7	42.8	25.5	9.1
		Normal	12.9	6.5	50.3	50.3	9.0
		Slow	12.4	5.1	41.3	24.5	7.5
	TKA	Fast	10.0	4.3	42.7	23.2	5.5
		Normal	10.0	4.9	49.3	27.2	6.5
		Slow	10.2	4.0	38.6	22.6	5.2
Toe out Angle (deg)	OA	Fast	23.9	15.7	65.8	71.8	21.3
		Normal	24.6	17.8	72.3	86.8	27.7
		Slow	27.4	16.7	60.9	72.4	26.1
	TKA	Fast	18.9	14.6	77.4	62.0	23.5
		Normal	20.8	15.9	76.5	105.0	23.4
		Slow	18.4	13.3	72.1	76.5	22.0
Step Time (s)	OA	Fast	0.5	0.1	14.7	0.3	0.1
		Normal	0.6	0.1	11.1	0.5	0.1
		Slow	0.7	0.1	13.6	0.6	0.1
	TKA	Fast	0.5	0.1	11.2	0.4	0.1
		Normal	0.6	0.1	9.9	0.4	0.1

		Slow	0.7	0.1	12.4	0.6	0.1
		Fast	0.9	0.1	13.7	0.6	0.2
	OA	Normal	1.1	0.1	10.0	0.7	0.2
Stride Time (s)		Slow	1.4	0.2	12.7	1.0	0.2
		Fast	1.0	0.1	10.2	0.6	0.1
	TKA	Normal	1.1	0.1	8.9	0.6	0.1
		Slow	1.4	0.2	11.7	1.0	0.2
		Fast	0.5	0.1	19.9	0.4	0.1
	OA	Normal	0.6	0.1	14.7	0.7	0.1
Stance Time (s)		Slow	0.8	0.1	16.7	0.8	0.2
		Fast	0.5	0.1	12.7	0.4	0.1
	TKA	Normal	0.6	0.1	11.2	0.5	0.1
		Slow	0.8	0.1	13.8	0.7	0.1
		Fast	0.4	0.1	11.7	0.3	0.1
	OA	Normal	0.5	0.1	11.4	0.4	0.1
Swing Time (s)		Slow	0.6	0.1	13.5	0.5	0.1
		Fast	0.4	0.0	9.8	0.3	0.1
	TKA	Normal	0.5	0.0	8.5	0.4	0.1
		Slow	0.6	0.1	11.9	0.5	0.1
		Fast	0.4	0.1	13.6	0.3	0.1
	OA	Normal	0.5	0.1	12.2	0.4	0.1
Single Support Time (s)		Slow	0.6	0.1	14.1	0.5	0.1
		Fast	0.4	0.0	9.7	0.2	0.1
	TKA	Normal	0.5	0.0	8.1	0.2	0.1
		Slow	0.6	0.1	11.8	0.5	0.1
		Fast	0.0	0.1	270.0	0.5	0.1
	OA	Normal	0.1	0.1	60.3	0.5	0.1
Double Support Time (s)		Slow	0.2	0.1	52.0	0.6	0.1
		Fast	0.1	0.0	70.3	0.3	0.1
	TKA	Normal	0.1	0.0	33.5	0.3	0.1
		Slow	0.3	0.1	30.0	0.6	0.1
		Fast	2.2	0.3	15.3	1.6	0.5
	OA	Normal	1.8	0.2	11.0	1.8	0.3
Cadence (1/s)		Slow	1.5	0.2	14.3	1.3	0.2
		Fast	2.1	0.3	12.6	1.9	0.2
	TKA	Normal	1.8	0.2	9.4	1.0	0.2

		Slow	1.4	0.2	12.8	1.2	0.2
		Fast	146.7	23.0	15.7	93.6	30.8
	OA	Normal	99.8	20.2	20.2	114.8	21.5
Speed (cm/s)		Slow	80.3	18.2	22.6	87.1	23.4
		Fast	139.7	26.1	18.7	122.0	38.5
	TKA	Normal	105.8	16.1	15.3	81.6	25.1
		Slow	77.5	14.2	18.3	74.7	17.3

3.4.2 Benchmarking Neural Network Architecture

MAE for stride length ranged from 2.9 ± 2.6 cm to 6.9 ± 3.2 cm for the validation set and 7.6 ± 6.1 cm to 11.9 ± 7.1 cm for the test set (Table 3.4). The CNN architecture proposed by Zrenner et al. yielded the lowest MAE for both the validation and test data sets, and the lowest ME, NAPE, and ME standard deviation for the test set, indicating negligible bias and low variance in the stride length predictions. Additionally, this network architecture included only 148,529 parameters which was smaller than the other networks, reducing the computational cost of training the network and preventing overfitting.

Table 3.4: Stride length prediction errors, mean error (ME), mean absolute error (MAE), and normalized absolute percentage error (NAPE) for multiple contemporary network architectures.

Models	ME ± STD (cm) Validation set	ME ± STD (cm) Test set	MAE ± STD (cm) Validation set	MAE ± STD (cm) Test set	NAPE ± STD (%) Validation set	NAPE ± STD (%) Test set	Number of Parameters
CNN (Hannink et al. 2016b)	0.5 ± 4.2	-2.2 ± 9.7	3.4 ± 2.9	8.2 ± 6.2	3 ± 2.5	7.2 ± 5.5	2,079,921
CNN (Zrenner et al. 2018)	**0.4 ± 3.7**	-2.4 ± 8.7	**2.9 ± 2.6**	**7.6 ± 6.1**	**2.5 ± 2.2**	**6.8 ± 5.5**	**148,529**

FCN (Wang et al. 2017)	-2.7 ± 3.9	-4.8 ± 9.1	8.4 ± 3.5	11.9 ± 7.1	7.3 ± 3	10.5 ± 6.3	277,121
ResNet (Wang et al. 2017)	0.5 ± 3.9	-1.9 ± 9.6	5.1 ± 3.2	9.1 ± 6.4	4.4 ± 2.8	8.1 ± 5.7	229,953
MLSTM-FCN (Karim et al. 2018)	1.0 ± 3.6	-1.2 ± 9.4	6.1 ± 3.1	9.5 ± 6.8	5.3 ± 2.7	8.3 ± 6	277,801
MALSTM-FCN (Karim et al. 2018)	1.0 ± 3.7	**-0.8 ± 9.0**	6.9 ± 3.2	10.3 ± 6.5	5.9 ± 2.7	9.1 ± 5.7	278,361
Hannink et al (Hannink et al. 2016b)	NA	-0.15 ± 6.1	NA	NA			2,079,921
Zrenner et al (Zrenner et al. 2018)	NA	2.5 ± 20.1	NA	15.3			

3.4.3 Optimal Sensor Combinations for Gait Characteristics

Across sensor combinations, network predictions for spatial gait characteristics were most accurate (lowest NAPEs) for step length ($7.6 \pm 6\% - 9.7 \pm 6.9\%$) and stride length ($7.1 \pm 5.7\% - 9.6 \pm 7.9\%$), followed by step width ($34.9 \pm 27.2\% - 40.9 \pm 32.9\%$) and toe-out angle ($73.7 \pm 50.9\% - 80.6 \pm 53.9\%$) (Figure 3.3). For temporal parameters, most accurate predictions were for step time ($3.1 \pm 2.9\% - 3.5 \pm 3.7\%$), stride time ($2.1 \pm 2.3\% - 2.6 \pm 3\%$), stance time ($3.5 \pm 3.5\% - 4.8 \pm 4.2\%$), and swing time ($4.6 \pm 4.1\% - 5.6 \pm 4.8\%$). Prediction errors increased for single support time ($5.2 \pm 4.4\% - 6.6 \pm 5.3\%$), and double support time ($22.6 \pm 18.1\% - 28 \pm 23.1\%$). For general parameters, cadence was predicted with the highest accuracy ($3.2 \pm 3.7\% - 4.1 \pm 4.6\%$) followed by speed ($6.4 \pm 5.2\% - 9.6 \pm 8.8\%$).
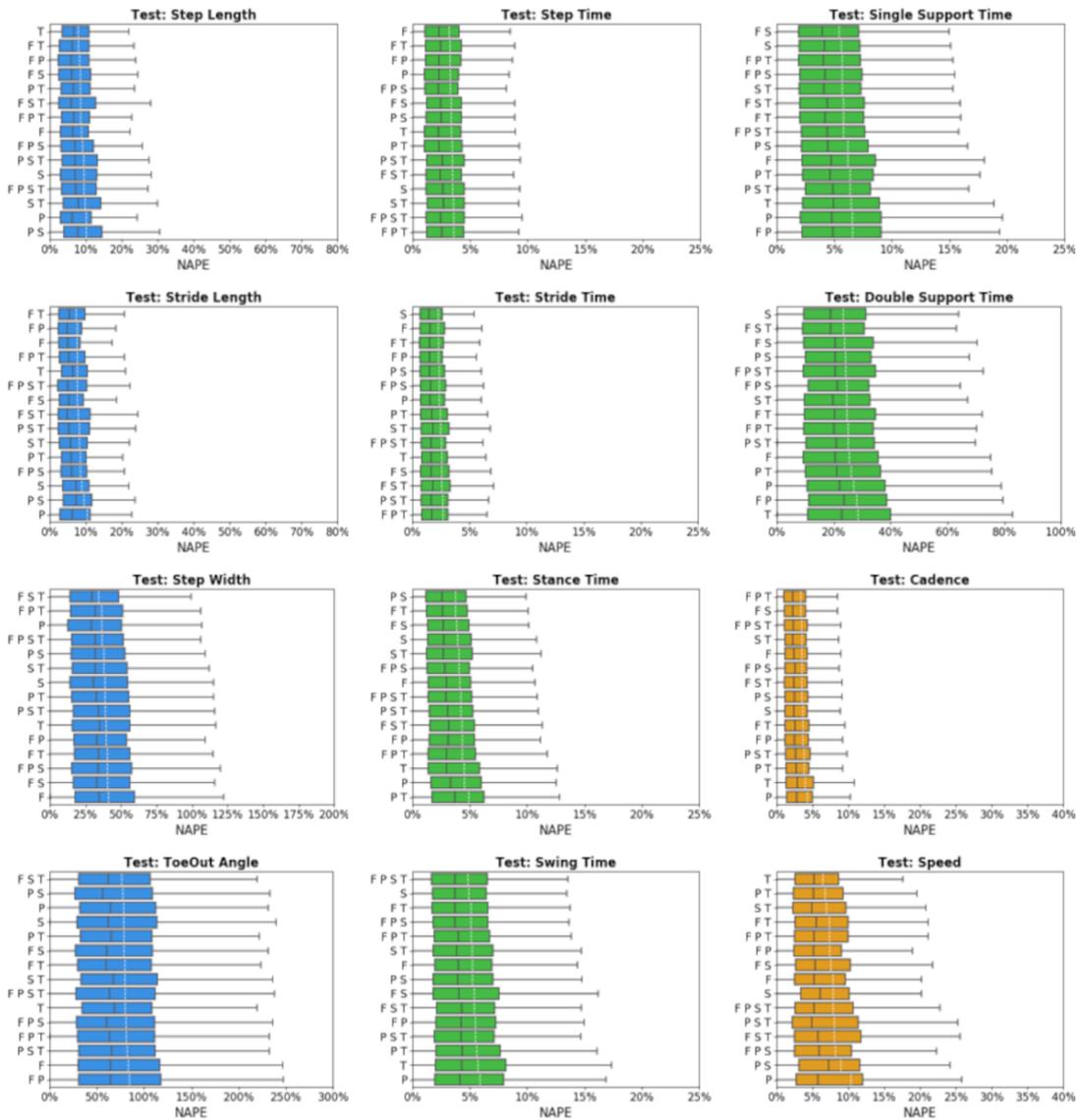
Figure 3.3: Normalized absolute percentage error (NAPE) of neural network for spatial (blue), temporal (green), and general (orange) gait parameters with various sensors configurations in the test set. Whiskers indicate 25% and 75% quartiles. For each gait parameter, sensor configurations are listed in order of increasing NAPE.

Predictive accuracy was not equivalent between the OA and TKA cohorts, with generally larger prediction errors for the OA cohort (Figure 3.4). The OA cohort had larger mean (19.0%) and median (6.6%) NAPE across all sensor combinations and STGPs

44

compared to TKA (mean NAPE = 14.7%, median NAPE = 4.6%). Fast walking also resulted in lower predictive accuracy relative to normal and slow walking. The mean and median NAPEs for fast walking were 17.7% and 6.22%, for normal walking were 15.8% and 4.8%, and for slow walking were 15.8% and 5.4% (Figure 3.5).
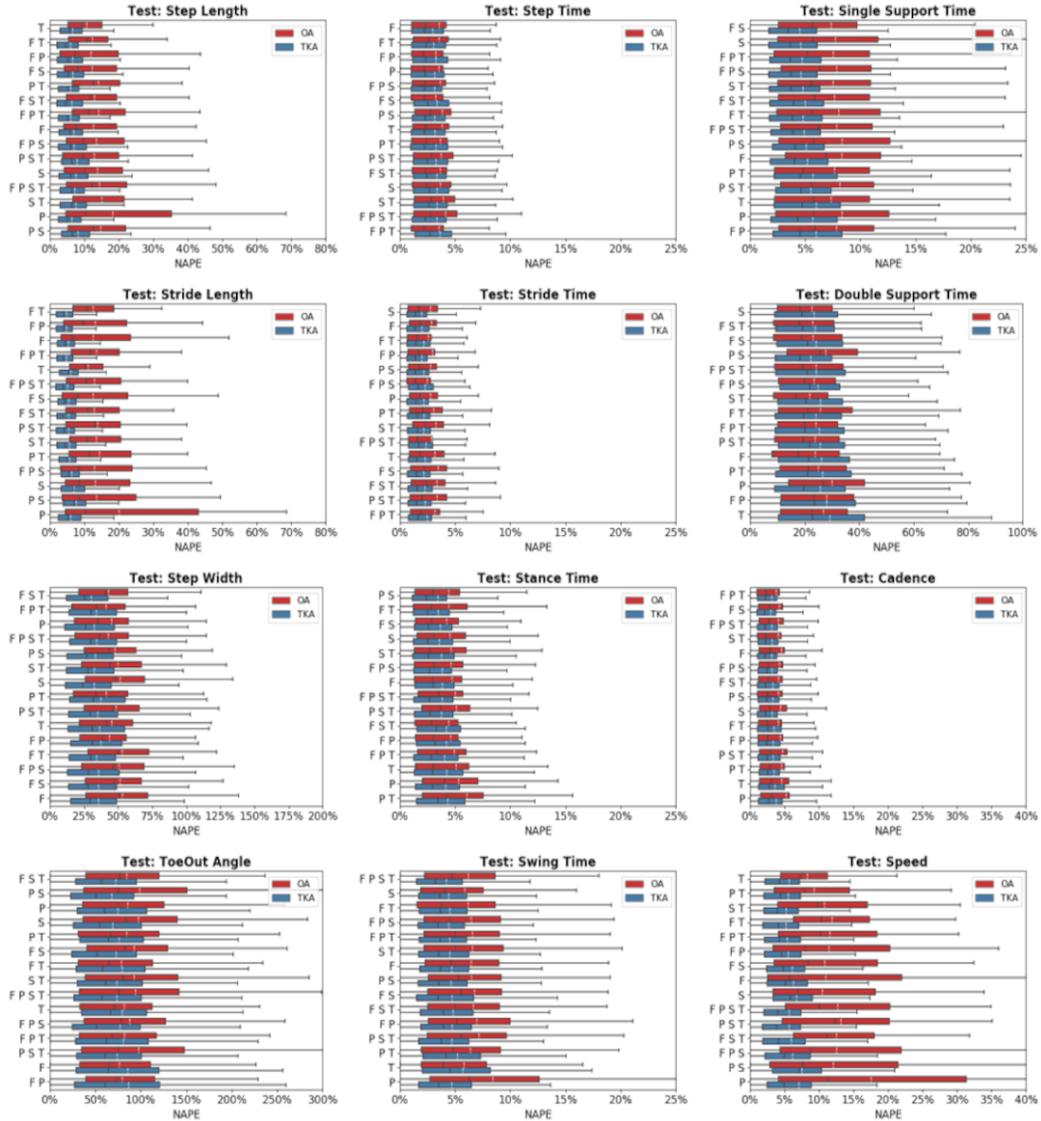
Figure 3.4: Normalized absolute percentage error (NAPE) of neural network predictions for all gait parameters and various sensors configurations grouped by subject cohort (OA and TKA) in the test set. For each gait parameter, sensor configurations are listed in order of increasing NAPE.
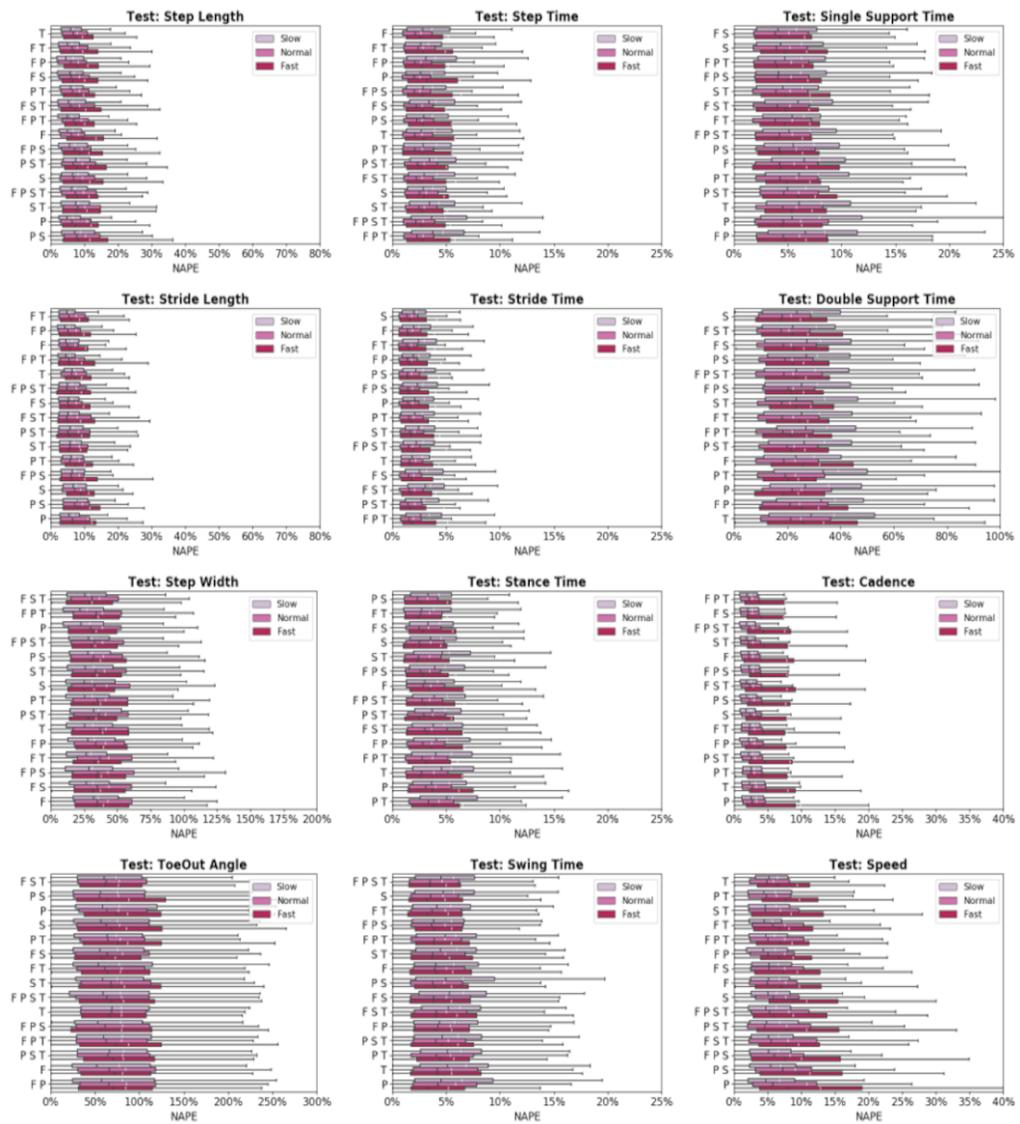
Figure 3.5: Normalized absolute percentage error (NAPE) of neural network predictions for all gait parameters and various sensors configurations grouped by gait pace (slow, self-selected, fast) in the test set. For each gait parameter, sensor configurations are listed in order of increasing NAPE.

None of the sensor combinations consistently yielded the highest prediction accuracy for all variables. Sensor combinations were ranked based on NAPE for each gait parameter (Table 3.5). Overall, the feet-thigh (F T) configuration had the best average rank (5.1), followed by the feet-shank (F S, 6.2), and shank (S, 6.3) sensor combinations. The shank

sensor combination consistently yielded the highest accuracy for temporal characteristics, ranking first or second for four of six temporal parameters. By contract, the pelvis (P) and pelvis-shank-thigh (P S T) sensor combinations consistently ranked among the least accurate with average ranks of 11 and 10.9, respectively.

Table 3.5: Sensor combinations ranking based on mean NAPE.

	F	P	F P	S	F S	P S	F P S	T	F T	P T	F P T	S T	F S T	P S T	F P S T
Step Length	8	14	3	11	4	15	9	1	2	5	7	13	6	10	12
Stride Length	3	15	2	13	7	14	12	5	1	11	4	10	8	9	6
Step Width	15	3	11	7	14	5	13	10	12	8	2	6	1	9	4
Toe Out Angle	14	3	15	4	6	2	11	10	7	5	12	8	1	13	9
Step Time	1	4	3	12	6	7	5	8	2	9	15	13	11	10	14
Stride Time	2	7	4	1	12	5	6	11	3	8	15	9	13	14	10
Stance Time	7	14	11	4	3	1	6	13	2	15	12	5	10	9	8
Swing Time	7	15	11	2	9	8	4	14	3	13	5	6	10	12	1
Single Support Time	10	14	15	2	1	9	4	13	7	11	3	5	6	12	8
Double Support Time	11	13	14	1	3	4	6	15	8	12	9	7	2	10	5
Cadence	5	15	11	9	2	8	6	14	10	13	1	4	7	12	3
Speed	8	15	6	9	7	14	13	1	4	2	5	3	12	11	10
Average Spatial	10.0	8.8	7.8	8.8	7.8	9.0	11.3	6.5	5.5	7.3	6.3	9.3	4.0	10.3	7.8
Average Temporal	6.3	11.2	9.7	3.7	5.7	5.7	5.2	12.3	4.2	11.3	9.8	7.5	8.7	11.2	7.7
Average General	7.8	13.0	7.4	7.6	6.6	10.8	10.5	5.3	4.5	5.7	6.3	5.2	9.6	11.0	8.8
Average	7.6	11.0	8.8	6.3	6.2	7.7	7.9	9.6	5.1	9.3	7.5	7.4	7.3	10.9	7.5

The Friedman test indicated statistically significant differences (p = 0.001) between sensor combinations. Multiple pairwise comparisons based on Friedman ranking are displayed as homogenous subsets in Table 6. Similar to the mean NAPE ranking, feet-thigh and feet-shank sensor combinations ranked first and second in Friedman ranking. There was not a statistically significant difference between feet-thigh and feet-shank (adjusted p-value = 0.077) while there was a statistically significant difference between feet-thigh and

the rest of the sensor combinations (adjusted p-value = 0.00). The pelvis sensor had the lowest accuracy with a significant difference compared to the other homogenous subsets of sensor combinations.

Table 3.6: Homogeneous subsets based on Freidman ranking and asymptotic significances (0.05).

	Subset						
	1	2	3	4	5	6	7
F T	7.637						
F S	7.759	7.759					
F P T		7.792					
F S T		7.818	7.818				
F P S T			7.887	7.887			
S			7.911	7.911			
F				7.946	7.946		
S T				7.952	7.952		
F P S				7.987	7.987		
F P					8.036		
P T						8.181	
T						8.200	
P S						8.229	
P S T						8.280	
P							8.386
Test Statistic	6.519	2.018	3.228	5.417	9.667	3.830	.
Sig. (2-sided test)	.011	.365	.199	.247	.022	.280	.
Adjusted Sig. (2-sided test)	.077	.896	.671	.573	.079	.709	.

48

3.5 Discussion

The primary outcome of this study was the development of a robust deep-learning framework to predict diagnostic gait parameters for subjects with OA and TKA and investigate various sensor combinations on prediction accuracy. A simple ensemble deep neural network with two layers of 1D-CNNs demonstrated robust performance in predicting each STGP compared to more complex networks. A design of experiment conducted on 15 combinations of sensors and locations for different patient populations and gait paces revealed how the prediction accuracy of STGPs can change over different conditions and identification of an optimal sensor combination might be challenging. Overall, feet sensors combined with either shank or thigh sensors produced the highest accuracy for most STGPs and the isolated pelvis sensor showed the lowest accuracy.

The CNN architecture proposed by Zrenner et al. resulted in the lowest MAE and the lowest standard deviations for both the validation and test subject datasets with errors of 2.9 ± 2.6 cm and 7.6 ± 6.1 cm, respectively (Zrenner et al. 2018). The CNN and ResNet model proposed by Hannik et al. and Wang et al. had the second and third lowest mean absolute errors of 8.2 ± 6.2 cm and 9.1 ± 6.4 cm (Wang et al. 2017). Both Hannik et al. and Zrenner et al. published the mean and standard deviation of their models' predictive error for stride length using unique datasets, enabling a direct comparison with our results (Hannink et al. 2016b; Zrenner et al. 2018). Hannick et al. predicted stride length based on more than 1300 strides from 101 geriatric subjects, with a mean error of -0.15 ± 6.09 cm compared to our error of -2.2 ± 9.7 cm using the same network architecture. They used a larger number of subjects (n = 99) compared to our study (n = 29) which gives more unique

data points for the network to train on. We induced additional variability in our dataset by asking subjects to walk at three different paces, all determined by the subjects. However, since this additional variability was mainly within-subject and may have a large amount of replication, it resulted in a slightly larger mean error compared to Hannink et al. In general, given large standard deviations in both studies, this difference was trivial. In this context, while our dataset had considerable variability, it likely had less variability than in the running dataset employed by Zrenner et al., which reported a mean predictive error in stride length of 2.5 ± 20.1 cm and a mean absolute error of 15.3 cm. The robustness of these CNN architectures for prediction of stride length point to the validity of using deep learning for this application, but also suggests that prediction accuracy is reduced when variability in the dataset is increased.

Direct comparisons of prediction accuracy between the current study and previous studies across all the STGPs are difficult due to differences in subject characteristics, dataset size, and experimental procedures. Comparable reported results from Hannick et al. for geriatric subjects, from Zrenner et al. for runners, and from Carcreff et al. for youths with cerebral palsy for sensor combinations and gait parameters are compiled in Table 3.7. Specifically, when comparing spatial parameter predictions using feet sensors, our results were within the range reported by previous studies. However, our results showed a larger mean error in prediction of stride length and step width compared to Hannink et al. that could be attributed to the larger number of subjects in Hannink et al. ($n = 101$) compared to our study. Diseases that induce pathologic movements, like OA, inherently increase the variability in gait parameters. The accuracy in prediction of the TKA group was higher

than the OA group. The NAPE for OA was 19.0% and for TKA was 14.7%. When accounting for this limitation, our errors and standard deviations were comparable to previously reported results.

The neural networks trained on all sensor combinations predicted spatial, temporal, and general parameters with varying levels of accuracy. The NAPE averaged across all sensor combinations, for step length, stride length, step width, and toe-out angle were 8.6 ± 0.7, 7.8 ± 0.7, 38.5 ± 1.8, $77 \pm 2\%$, respectively. The increased predictive error for step width and toe-out angle was likely associated with the smaller mean movements for those parameters, reducing the signal-to-noise ratio compared to the larger sagittal plane motions. For temporal parameters, the NAPE ranged from $2.3 \pm 0.1\%$ for stride time to $24.9 \pm 1.5\%$ for double support time. For the general parameters, the NAPE was 3.5 ± 0.2 and $7.5 \pm 0.8\%$ for cadence and speed, respectively. Descriptive statistical analysis on STGPs in our dataset revealed that neural network predictions were more accurate for the parameters with a lower coefficient of variation (CV). CV was defined as the ratio of the standard deviation to the mean which is an indicator of the dispersion of a probability distribution of data (SRJ and Everitt 1999). This was evident in the lower prediction accuracy observed for step width, toe-out angle, double support time, and speed with larger CVs compared to other parameters (Table 3.3).

Table 3.7: Deep-learning accuracy comparison with previous studies for (a) spatial parameters, (b) general, and (c) temporal parameters.

(a)	Spatial ME ± STD			
	Step Length (cm)	Stride Length (cm)	Step Width (cm)	Toe-Out Angle (deg)

Feet	Our Results	-1.7 ± 5.2	-3.0 ± 8.7	1.1 ± 5.1	-3.2 ± 15.8
	Hannink	NA	-0.15 ± 6.09	-0.09 ± 4.22	NA
	Zernner	NA	2.5 ± 20.1	NA	NA
	Carfcreff	NA	2.5 ± 3.7	NA	NA
Shank Thigh	Our Result	-0.6 ± 5.6	0.4 ± 9.7	0.85 ± 4.6	-3.7 ± 15.2
	Carfcreff	NA	7.5 ± 6.9	NA	NA
Average of all sensors	Our Test	-0.5 ± 0.6	-1.1 ± 1.3	0.6 ± 0.4	-3.5 ± 0.8

(b)		General ME ± STD	
		Cadence (1/s)	Speed (cm/s)
Feet	Our Results	0.02 ± 0.1	-1.9 ± 8.3
	Hannink	NA	NA
	Zernner	NA	0.05 ± 0.28
	Carfcreff	NA	0.3 ± 4.5
Shank Thigh	Our Result	0.01 ± 0.09	-0.45 ± 8.2
	Carfcreff	NA	7.3 ± 6.7
Average of all sensors	Our Test	0.04 ± 0.00	0.0 ± 1.0

(c)		Temporal ME ± STD		
		Stride Time (s)	Stance Time (s)	Swing Time (s)
Feet	Our Results	-0.01 ± 0.04	-0.01 ± 0.03	0.01 ± 0.03
	Hannink	-0.00 ± 0.07	-0.00 ± 0.07	0.00 ± 0.05
	Carfcreff	0.00 ± 0.02	NA	NA
Shank Thigh	Our Results	0.00 ± 0.03	-0.01 ± 0.04	0.00 ± 0.03
	Carfcreff	0.00 ± 0.02	NA	NA
Average of all sensors	Our Results	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

The differences in predicted accuracy for OA versus TKA groups was multifactorial. First, there was more variability in the gait of OA subjects due to their pathology which makes it harder to predict certain STGPs. This higher variability for OA subjects was expressed by higher standard deviations and coefficients of variation for all gait parameters except toe-out angle, stride time, and cadence (Table 3.3). Second, the accuracy in prediction of STGPs was slightly higher at normal and slow walking compared to fast walking with mean NAPE of 15.8% for normal and slow, and 17.6% for fast walking. This aligns with findings by Zrenner et al. that indicated increasing speed could negatively impact predictive accuracy due to higher variability at fast walking (Table 3.7). Stressing the OA group with higher demand walking at a fast pace resulted in even greater variability and decreased predictive accuracy.

Perhaps most important, one of the randomly selected OA test subjects (subject S21, see Appendix A Figure .A1 and Figure .A2), walked with the shortest step length, shortest stride length, largest step width, and slowest speed among all subjects in the study, making this subject an outlier. Since our sample size was small, the impact of a single outlier was amplified and negatively affected prediction results. Jensen–Shannon divergence, which measures the similarity between two probability distributions, showed a larger divergence for subject S21 compared to the other two subjects in the same fold (S19 and S27). The divergences of step length, stride length, and step width for subject S21 were 5.67, 7.01, and 5.96 while for subject S19 and S27 the divergences were 0.18, 0.17, 0.10 and 0.38, 0.59, 0.67, respectively (see Appendix A Table A1). The divergence of S21 from the distribution of subjects used to train the CNNs resulted in poor performance, driving up

the reported error for the OA cohort. Removing subject S21 from the test set reduced the mean and median NAPE from 19.0% and 6.6% to 17.3% and 4.8% which is comparable to the TKA group. Subject S21 had severe knee OA of the right knee which caused pain during activities of daily living and manifested in a noticeable limp on the affected limb compared to the other subjects in the OA cohort. Investigation of NAPE from the validation set revealed almost equal performance on both knee groups with mean and median NAPE of 7.7% and 2.9% for OA and 7.4% and 2.8% for TKA. The impact of subject S21 in the test set is an example of how CNNs result in poor performance when faced with data that are outside the distribution of the training data, which is one of the main challenges in the use of machine-learning models for real-world applications. Hence, gaining intuition on training data set completeness is important prior to interpreting prediction accuracy. Out-of-distribution detection (DeVries and Taylor 2018; Liang et al. 2017) has recently been recognized as an important area for developing trustworthy machine learning (DeVries and Taylor 2018; Liang et al. 2017)and will be continually addressed in this work as patient numbers increase.

Our statistical analysis indicated statistically significant differences in accuracy between various sensor combinations tested across all conditions. The F-T combination was the highest-ranked sensor combination based on a Friedman test, showing a significant improvement in accuracy with respect to every other sensor combination except the F-S. It should be noted that although statistically significant, differences between the most and least accurate sensor combinations were small. The best sensor combination based on mean NAPE was F-T-S (15.25%) while the worst sensor combination was F-P (16.65%).

Similarly, the Friedman test indicated the sensor combinations of F-T, F-S, and F-P-T were the top three ranked sensor combinations TKA subjects while T, F-S, and F-T were the top three ranked sensor combinations for OA subjects (see Appendix A Table A 2a,b). The F-T and F-S were the common sensor combinations suitable for both OA and TKA groups. In addition, the F-T combination was also among the top three for slow, normal, and fast walking paces (see Appendix 3.A Table 3.A2 c,e). As noted earlier, while the F-T sensor combination proved to be statistically better than other combinations, a 2–5% improvement in overall STGP prediction accuracy may be impactful during certain clinical applications. For instance, given the small difference in stride length at the normal pace between OA and TKA groups (~3 cm) higher accuracy predictions may be necessary for diagnostic purposes. However, higher accuracy may not be important for parameters with large differences between patient groups. This is an advantage of data driven modeling compared to other algorithm-based techniques in the prediction of STGPs. If the accuracy of STGPs is not largely impacted by sensor combination, there is freedom to design patient monitoring systems for specific patient groups based on other factors, such as cost and patient compliance. Feet sensors were necessary for stride segmentation during gait which is an input for the trained models. Therefore, including feet sensors is imperative for using a data-driven approach. Testing these sensor combinations on more complex tasks such as climbing stairs, sit-to-stand, and evaluating other joint kinematic and kinetic parameters would be necessary to clarify the value of using certain sensor combinations.

There are limitations to this study that should be considered. This study focused on gait to demonstrate the ability to predict STGPs from IMU data. In the OA population, other

activities of daily living that place a greater demand on the subject will likely provide additional clinical value. The methods demonstrated in this study can be extended to predict analogous spatial temporal parameters for activities that include stair ascent/descent, sit-to-stand, and other high-demand activities. This study was also limited in the number of subjects that were included. This study demonstrated acceptable accuracy with 3778 segmented and labeled strides from the 29 subjects. Increasing the number of subjects and labeled strides will improve the predictive accuracy. Like other data-driven approaches, the trained network described in this study are only suitable for the selected population. There are also practical limitations to deploying our algorithm to a large patient population outside of a laboratory environment, including variability in sensor placement, reduced signal quality from low-cost IMUs, soft-tissue artifacts for high body mass index patients, and identification of patients with gait parameters outside the training data set. In order to implement this workflow for other populations with movement impairments that would benefit from patient monitoring, such as patients with cerebral palsy or stroke, the algorithm would need to be re-trained with inclusion of data from these populations. However, with this initial model architecture defined and the trained, a transfer-learning approach could be used on other populations to drastically reduce training time and the need for high volumes of data.

3.6 Conclusions

This study demonstrated that a deep-learning, data-driven approach was able to predict spatial temporal gait characteristics of OA and TKA patients based on signals from IMU sensors. Using a comprehensive analysis of various sensor combinations and their

sensitivity to STGPs, patient population, and walking pace, our results showed that deep learning can overcome the dependency on sensor location that hinders the design of patient monitoring systems and negatively impacts patient compliance. Additionally, we demonstrated the importance sufficient variability in training and test data as a critical factor in the performance of DL models, especially for clinically relevant data with small sample sizes. A system that is able to leverage data streams from wearable sensors to produce real-time monitoring of STGPs in OA and TKA patients has the ability to improve clinical care and patient quality of life.

CHAPTER 4: THE USE OF SYNTHETIC IMU SIGNALS IN THE TRAINING OF DEEP LEARNING MODELS SIGNIFICANTLY IMPROVES THE ACCURACY OF JOINT KINEMATIC PREDICTIONS

4.1 Abstract

Gait analysis based on inertial sensors has become an effective method of quantifying movement mechanics, such as joint kinematics and kinetics. Machine learning techniques are used to reliably predict joint mechanics directly from streams of IMU signals for various activities. These data-driven models require comprehensive and representative training datasets to be generalizable across the movement variability seen in the population at large. Bottlenecks in model development frequently occur due to the lack of sufficient training data and the significant time and resources necessary to acquire these datasets. Reliable methods to generate synthetic biomechanical training data could streamline model development and potentially improve model performance. In this study, we developed a methodology to generate synthetic kinematics and the associated predicted IMU signals using open source musculoskeletal modeling software. These synthetic data were used to train neural networks to predict three degree-of-freedom joint rotations at the hip and knee during gait either in lieu of or along with previously measured experimental gait data. The accuracy of the models' kinematic predictions was assessed using experimentally measured IMU signals and gait kinematics. Models trained using the synthetic data out-

performed models using only the experimental data in five of the six rotational degrees of freedom at the hip and knee. On average, root mean square errors in joint angle predictions were improved by 38% at the hip (synthetic data RMSE: 2.3°, measured data RMSE: 4.5°) and 11% at the knee (synthetic data RMSE: 2.9°, measured data RMSE: 3.3°), when models trained solely on synthetic data were compared to measured data. When models were trained on both measured and synthetic data, root mean square errors were reduced by 54% at the hip (measured + synthetic data RMSE: 1.9°) and 45% at the knee (measured + synthetic data RMSE: 1.7°), compared to measured data alone. These findings enable future model development for different activities of clinical significance without the burden of generating large quantities of gait lab data for model training, streamlining model development, and ultimately improving model performance.

4.2 Introduction

Gait analysis and musculoskeletal modeling (MSM) are commonly used to quantify movement mechanics, providing insights into the diagnosis, treatment, and rehabilitation of movement disorders (Clary et al. 2013; Sartori et al. 2016a). Using the current gold-standard passive-marker motion capture (MOCAP) systems, detailed kinematic measurements are time consuming, constrained to laboratory environments, and require technical expertise to generate reliable data. Wearable inertial measurement units (IMUs) enable biomechanical measurements without many of the logistical constraints of traditional techniques by translating multiple streams of IMU data into an accurate measurement of joint mechanics. However, establishing reliable clinical metrics of pathological movement with the use of IMUs remains a major hurdle.

Early IMU-based methods for measuring lower limb kinematics integrate the rotational velocity and linear acceleration data from each limb segment, coupled with orientation data from the magnetometer, to make estimations about limb segment positions and orientations (Picerno 2017; Weygers et al. 2020). These methods are prone to errors imparted by noise, drift, and other inaccuracies in IMU signals. More recently, the accuracy of IMU-based kinematic measurements has been improved by integration with MSMs and optimization algorithms to impose realistic joint constraints to the estimated movements (Dorschky et al. 2019; Karatsidis et al. 2018, 2019; Konrath et al. 2019; Seel et al. 2014). These methodologies, however, require nontrivial computational resources making them less suitable for real-time applications with instantaneous feedback (Dorschky et al. 2019; Gholami et al. 2020; Halilaj et al. 2018).

Despite the intensive computational resources necessary for training machine learning algorithms, trained models can be deployed with minimal processor power to generate instantaneous kinematic and kinetic predictions. These techniques include neural networks (NNs) to estimate ground reaction forces for gait, running and jumping(Hendry et al. 2020; Johnson et al. 2019, 2020; Komaris et al. 2019; Wouda et al. 2018), and lower limb joint kinematics and kinetics (Argent et al. 2019; Dorschky et al. 2019; Gholami et al. 2020; Halilaj et al. 2018; Lim et al. 2019; Stetter et al. 2020; Zaroug et al. 2020). The accuracy of these algorithms relies on large and representative biomechanics training datasets that are frequently expensive and time consuming to collect. To expand the availability of training data, researchers are leveraging artificially generated data to improve model prediction accuracy and reliability.

The most common technique to generate artificial IMU data in movement analysis is to leverage existing passive-marker motion capture datasets to calculate simulated IMU data based on marker trajectories and accelerations (Brunner et al. 2015; Johnson et al. 2019, 2020; Mundt et al. 2019; Young et al. 2014; Zimmermann et al. 2018). For the purposes of this paper, artificial IMU data generated using this technique will be referred to as "simulated IMU" data. Using this technique, Mundt et al. (Mundt et al. 2020b) combined simulated IMU data from an archived MOCAP database with experimentally measured IMU data on a smaller subject cohort to predict lower limb kinematics and kinetics during gait. The inclusion of the simulated data in the training set reduced the root mean square error in joint kinematic estimates from 4.8° to 4.3° but did not improve joint kinetics predictions. The authors attributed the modest prediction improvements to inaccuracies in the simulated IMU data, specifically the lack of soft tissue-induced vibrations. One limitation of this technique is that observations are confined to movements measured in the lab and potentially do not span the variability present in the population at large.

Dorschky et al. (Dorschky et al. 2020) combined measured IMU data from subjects during walking and running with artificial IMU data generated from complementary MSMs. The authors applied perturbations to the MSM's joint angles, ground reaction forces, and speeds based on random sampling from the experimental measures to generate synthetic IMU data for movements not observed experimentally. The constraints of the MSM and corresponding optimal control algorithm ensured the perturbations resulted in physically realistic joint mechanics. For the purposes of this paper, artificial IMU data

generated for movements beyond those observed experimentally will be referred to as "synthetic IMU" data. Similar to Mundt et al., the addition of the synthetic data improved kinematic predictions at the hip, knee, and ankle but had mixed effects on the joint moments and ground reaction forces. In this way, both simulated IMU data (IMU data generated from existing MoCap data) and synthetic IMU data (IMU data generated from artificial kinematics not measured in the lab) are useful for expanding model training sets, but synthetic IMU data enables expansion of the training set to uncommon movements that are difficult to measure in the lab.

Numerical techniques to supplement existing optical tracking data with simulated IMU data or augmentation techniques to expand existing datasets with unique synthetic observations have both proven effective at enhancing kinematic predictions from machine learning algorithms. However, both techniques rely on intensive gait lab data collections, limiting widespread accessibility. In contrast, synthetic data can be generated with only a few representative gait lab measurements of an activity to establish the general kinematic patterns of the movement. It has yet to be demonstrated if machine learning models can achieve the necessary accuracy when trained exclusively with synthetic data. In this study, we aimed to develop a musculoskeletal modeling framework to augment and expand an existing dataset of gait kinematics, then use those synthetic data to train neural networks to predict 3-D joint angles from experimentally measured IMU data during gait. We hypothesize that (1) introducing synthetic IMU data into the training dataset will significantly improve the kinematics predictions and (2) models trained exclusively on

synthetic data will perform equivalent to models trained using experimentally measured IMU data.

4.3. Materials and Methods

4.3.1 MU Measurement and Simulation Workflow Overview

In this study, we trained recurrent neural networks to predict three-dimensional hip and knee kinematics during gait using either experimentally measured IMU data, synthetically generated IMU data, or a combination of experimental and synthetic IMU data. First, combined IMU and motion capture data were collected from 30 subjects during multiple gait trials at various speeds (Figure 4.1). Next, the subjects' lower limb hip and knee kinematics were calculated from the experimental marker position data, then augmented to generate synthetic IMU data via a musculoskeletal modeling workflow in OpenSim (Delp et al. 2007). Finally, the original measured kinematic data, the synthetic kinematic data, and the combined measured and synthetic data were used to train the recurrent neural networks to estimate three-dimensional hip and knee kinematics during gait from IMU data. The detailed experimental methods can be found below.
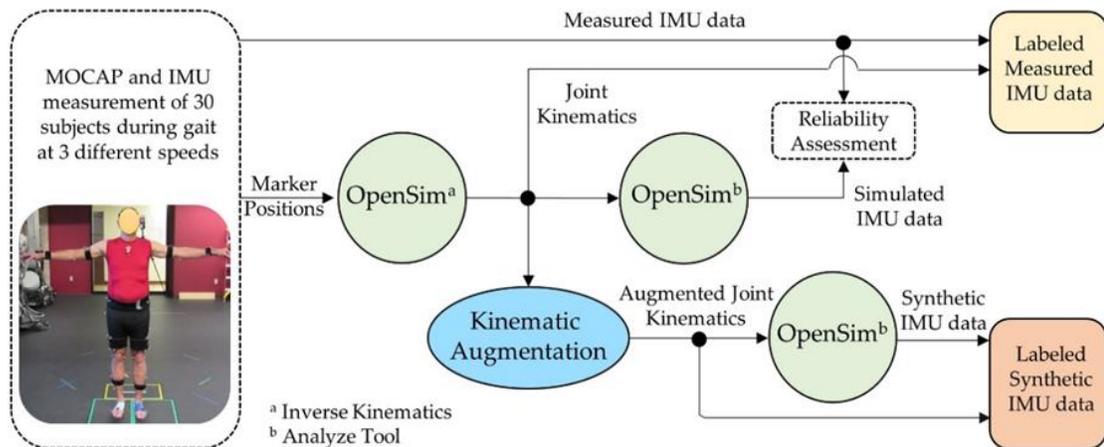
Figure 4.1: Overview of workflow for generating IMU signals labeled with joint kinematics. Knee and hip kinematic were calcu-lated from measured marker positions using the Inverse Kinematic tool in OpenSim. The OpenSim analysis tool was used to generate simulated IMU data from the experimental kinematic and synthetic data from augmented joint kin-ematics. Simulated IMU signals from the experimental kinematics were compared with the measured IMU data to de-termine the reliability of the IMU simulation process.

4.3.2 Experimental Data Collection

In total, 30 subjects, including 13 subjects with OA (age = 63 ± 6, weight = 76 ± 14 kg, height = 165 ± 13 cm, 6 females and 8 males) and 17 subjects with total knee arthroplasty (age = 68 ± 5, weight = 76 ± 14 kg, height = 163 ± 13 cm, 13 females and 4 males), participated in the study as part of a larger investigation. All participants signed a consent form prior to the experiment with IRB approval (IRB# 1328728). All biomechanical measurements were carried out in the same lab setting. Subjects were outfitted with 71 reflective markers on anatomical landmarks and 17 research-grade IMUs on various limb segments and the trunk. Only IMUs located on the pelvis, left thigh, left shank, and left foot were used in this analysis (Renani et al. 2020). Thigh and shank IMUs were attached to rigid 4-marker clusters used to track the relative orientations of the IMUs, while markers

placed directly on the IMUs were used to track IMU displacements. The relative orientation of the foot IMU was tracked using markers on the medial and lateral malleoli in addition to a marker directly on the foot IMU. Similarly, the relative orientation of the pelvis IMU was tracked using markers placed on the posterior superior iliac crests and a marker on the pelvis IMU.

Subjects performed 15 trials of a 5 m walking task at three different speeds: self-selected, slow, and fast. During the walking trials, synchronized data were collected from a 13 camera Vicon motion capture system (Centennial, CO), 4 Bertec force platforms (Columbus, OH), and the IMUs (Xsens, Enschede, The Netherlands) (Figure 4.2a). The sampling frequency of force data, MOCAP, and IMUs (acceleration and angular velocity) were 1000 Hz, 100 Hz, and 40 Hz, respectively. The IMUs used in this study leveraged on board data processing to reduce noise and drift in the signals. IMU data for each trial were upsampled using cubic interpolation to 100 Hz and filtered using a Butterworth low-pass filter with cutoff frequency of 6 Hz.
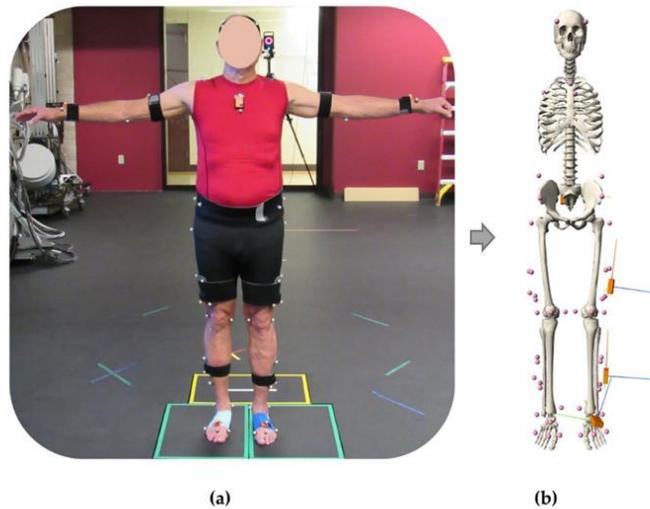
Figure 4.2: Subject fitted with reflective markers and inertial measurement units (a) and the corresponding musculoskeletal model representation with the virtual IMU sensors (b).

4.3.3 Musculoskeletal Modeling and IMU Simulation

Subject-specific musculoskeletal models were created for each study participant using a previously published workflow (Myers et al. 2018). Each model included 10 rigid body segments, 23 degrees of freedom, and 92 muscle actuators. The hip and knee joints were each modeled with three rotational degrees of freedom, and the ankle was modeled as a 1 degree of freedom hinge joint. Limb segments were scaled to match the optical markers from the experiment. An inverse kinematics analysis was conducted for each subject and each gait trial using OpenSim to obtain 3-D joint kinematics at the hip and knee (Delp et al. 2007).

Four virtual IMUs were placed on the pelvis, thigh, shank, and foot according to their experimentally measured locations and orientations via fixed joints to their respective limb segments (Figure 4.2b). These virtual IMUs were used in subsequent steps to generate

synthetic IMU signals. The angular velocities and linear accelerations of the model's rigid segments (pelvis, thigh, and shank) from the inverse kinematics analysis were used to calculate simulated IMU signals using the Analyze Tool in OpenSim. The angular velocity of each rigid segment in the global coordinate system was transformed through the local segment's anatomic reference frame and into the simulated IMU's local sensor-based coordinate system to align with the experimental measurements. To calculate the simulated IMU's acceleration, the second derivative of the position vector for the marker placed directly on the IMU was calculated in the global OpenSim coordinate system. These accelerations were transformed into the experimental IMU's global earth-fixed coordinate system for comparison to the experimental measurements. To assess the reliability of the simulated IMU data, the root mean square error (RMSE) and the Pearson correlation coefficient (r) between the experimentally measured IMU data and the simulated IMU data were calculated.

4.3.4 Kinematic Augmentation and Synthetic IMU Data Generation

Joint angles calculated from the measured data were segmented into individual gait cycles of the left lower limb using the heel marker resulting in 3943 unique strides from the 30 study participants. These joints angles were augmented using five different numerical techniques to introduce variation in both the magnitude of the joint angles (e.g., increased knee flexion during stance) and the timing of the gait events (e.g., shorter stance phase). These methods included magnitude offsets, magnitude warping, combinations of magnitude offsets and warping, time warping, and combinations of time warping and magnitude warping (Figure 3) (Tran and Choi 2020; Um et al. 2017; Wen et al. 2020).

Magnitude offsets were introduced by adding a random number from a normal distribution ($\mu = 0°$, $\sigma = 5°$) to all joint angles from a given trial. Magnitude warping was introduced by fitting a cubic spline to seven random numbers from a normal distribution ($\mu = 1$, $\sigma = 0.2$) that were uniformly spaced along the time domain of the input gait cycle. The cubic spline was evaluated at the time increments from the original trial's joint angle vectors to form a distortion vector. Joint angles augmented with magnitude warping were generated by multiplying corresponding elements of the distortion vector and the original joint angle vector. The same distortion vector was used on all joint angles from a given gait trial. Augmented joint angles with combined offset and warping were generated using these same methods by first applying the magnitude warping and subsequently applying the magnitude offset.

Time warping was introduced using a similar methodology by fitting a cubic spline to seven random numbers from a normal distribution ($\mu = 1$, $\sigma = 0.2$) that were spaced uniformly along the time domain of the input gait cycle. The cubic spline was evaluated at the time increments from the original trial's joint angle vectors, then the cumulative sum vector was calculated and divided by the length of the original joint angle vector to form a time distortion vector. Joint angles augmented with time distortion were generated by interpolating the original joint angles at the time values in the time distortion vector. Augmented joint angles with combined time warping and magnitude warping were generated by first applying the time warp to the joint angles and then applying the magnitude warp.

One set of augmented joint kinematics was generated for every gait trial using each of the five augmentation methods described above, resulting in a total of 19,715 sets of augmented joint angles from the original 3943 measured strides (5:1 ratio). Synthetic IMU data were calculated for each set of augmented joint kinematics by using the new joint angles to animate the associated patient-specific musculoskeletal model, in lieu of optical marker locations, using the workflow described above in OpenSim Analyze. This kinematics augmentation method introduced random variation into the dataset, and no controls were implemented to ensure the resulting kinematics were strictly physiological. Kinematic perturbations were selected from normal distributions with conservative standard deviations ($\sigma = 5°$ for angular offsets, 20% for magnitude warping, and 20% for time warping) that ensured the perturbations were similar to the measured kinematics. In this way, generation of the augmented kinematic data required no a priori knowledge of movement strategies, making the results more generalizable to other movements of interest and simpler to implement.

All experimentally measured IMU data and the corresponding joint angles were lowpass filtered with a second-order Butterworth filter using a cutoff frequency of 6 Hz (Choi et al. 2019; Gholami et al. 2020; Żuk and Pezowicz 2015). In addition, all datasets were zero padded to a length of 200, corresponding to the maximum length of any stride in the dataset.
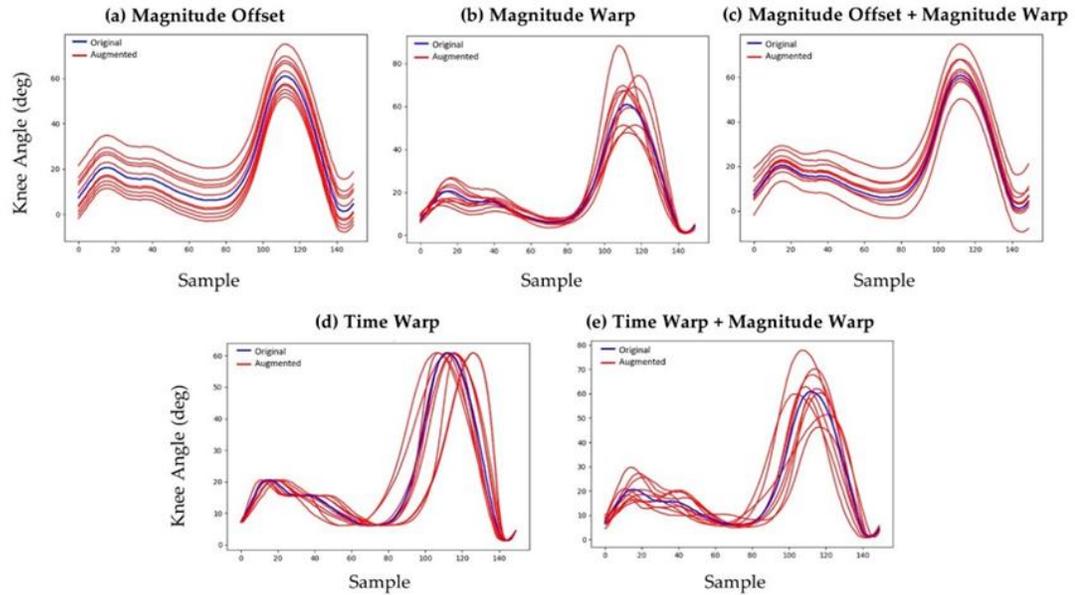
Figure 4.3: Various data augmentation methods used to generate synthetic kinematic data: (a) magnitude offset, (b) magnitude warping, (c) combined magnitude offset and magnitude warping, (d) time warping, and (e) combined time warping and magnitude warping.

4.3.5 Neural Network Model Architecture, Tuning, Training, and Evaluation

To facilitate neural network model development and testing, subjects from the experimental dataset were randomly assigned into training and test groups. The training dataset included the experimental measurements for all gait cycles from 27 subjects (3451 gait cycles). All experimental measurements from the remaining three subjects were reserved for the test set (492 gait cycles).

Two independent neural network models, one for knee kinematics and one for hip kinematics, were developed to predict joint angles from the corresponding IMU data. Both networks contained a bidirectional long short-term memory (BiLSTM) layer, followed by two fully connected layers. LSTM models are a specific class of recurrent neural networks particularly suited to time series data by addressing the vanishing gradient problem.

70

Specifically, LSTM is unique in that it uses feedback to remember long-term dependencies of the input data on output data with the use of the time domain. Unlike unidirectional LSTM models, which only consider information from the past, BiLSTM models also invert the time scale of the data to consider information from the future input, which may inform the present prediction and ultimately improve accuracy (Burton et al. 2021; Siami-Namini et al. 2019). A dropout of 0.5 was added prior to the final layer to avoid overfitting. The model input was a 200×24 matrix containing the three accelerations and three angular velocities from the pelvis, left thigh, left shank, and left foot IMUs. The model output was a 200×3 matrix containing the corresponding 3-D joint angles of the hip or knee (flexion–extension, adduction–abduction, and internal–external rotations) as a function of time. Model training was conducted using an adaptive learning rate optimization with a learning rate, beta-1, and beta-2 of 0.001, 0.9, respectively, with a total of 100 epochs (Kingma and Ba 2014). The size of each batch size was 50. The model development and training were conducted using PyTorch.

The neural network models' hyperparameters, including the number of BiLSTM layers and hidden sizes, were tuned via 5-fold cross-validation using only the experimentally measured MOCAP and IMU data in the training dataset (Raschka 2018). Specifically, the training set was subdivided into 5 sets with 5–6 subjects per set. In each fold, one set was reserved for validation, while the remaining four were used to train models with all combinations of hyperparameters. The prediction accuracies of these models and corresponding hyperparameters were evaluated on the validation set designated for that fold. Hyperparameters that resulted in the minimum average RMSE across all five folds

for each model were used in all subsequent model training and evaluation (Table 4.1). The optimal hip model had 1 layer with a hidden size of 32, while the optimal knee model had 1 layer with a hidden size of 128.

Table 4.1: Architecture of bidirectional long short-term memory neural networks with tuned hyperparameters.

Model	n-Layers Evaluated	Hidden Sizes Evaluated	Optimal n-Layer	Optimal hidden size
Hip-BiLSTM	[1, 2, 3, 4]	[16, 32, 64, 96, 128]	1	32
Knee-BiLSTM	[1, 2, 3, 4]	[16, 32, 64, 96, 128]	1	128

To investigate the influence of synthetic IMU and lower limb kinematic data on prediction accuracy, hip and knee neural network models were trained on three variations of the training dataset. The first variation included the experimentally measured IMU signals and associated kinematics for gait cycles from all 27 subjects included in the training cohort (3451 measured gait cycles). The second variation included only the synthetic IMU signals and associated kinematics generated from the subjects in the training cohort (17,255 synthetic gait cycles). The third dataset included the measured data from the training cohort and the corresponding synthetic data generated for those same subjects (3451 measured gait cycles and 17,255 synthetic gait cycles). All three sets of trained models were used to predict the lower limb joint angles for all trials of the three subjects assigned to the test cohort using the measured IMU data (492 measured gait cycles). The predictive accuracy of the models was quantified by calculating the RMSE, normalized-RMSE, and Pearson correlation coefficients between the predicted and measured joint angles from the test set. A multivariate analysis of variance (MANOVA) was performed

72

with the predicted RMSE for each of the six kinematic degrees of freedom (e.g., Knee Flex–Ext or Hip Ad–Ab) as the dependent variables and the training dataset type as the independent variable. Tukey's honest significant different (HSD) post hoc tests were performed to determine which kinematic degrees of freedom demonstrated statistically significant prediction improvements with each training dataset ($p < 0.05$).

4.4 Results

4.4.1 Simulated IMU Accuracy

The average RMSE between the measured IMU data and simulated IMU based on marker trajectories across all sensors for angular velocities was 0.56 rad/s (ranging from 0.33 to 1.02) with correlation coefficients ranging from 0.29 in the pelvis sensor's y-axis to 0.98 in the foot sensor's y-axis (Table 4.2, Figure 4.4). Similarly, the average RMSE for accelerations was 1.43 m/s2 (ranging from 0.62 to 2.46) with correlation coefficients ranging from 0.75 in the thigh sensor's y-axis to 0.96 in the shank sensor's x-axis. IMU predictions for free accelerations were generally more accurate than angular velocities with average correlations coefficients of 0.86, compared to 0.71, respectively. Predictions for the pelvis IMU were consistently worse than predictions for the other segments, particularly for the pelvis rotational velocities (mean $r = 0.47$).

Table 4.2: Pearson correlation coefficient (r) and root mean square error (RMSE) and of angular velocity and acceleration between simulated and measured IMU data across all subjects and trials.

Segment	IMU DoF	Angular Velocity (rad/s)			Acceleration (m/s²)		
		r (Mean ± Std)	RMSE (Mean ± Std)	nRMSE (Mean ± Std)	r (Mean ± Std)	RMSE (Mean ± Std)	nRMSE (Mean ± Std)
Pelvis	x	0.62 ± 0.15	0.40 ± 0.17	19.47 ± 4.90	0.88 ± 0.10	0.65 ± 0.30	11.03 ± 3.79
	y	0.29 ± 0.24	0.36 ± 0.15	32.28 ± 11.00	0.79 ± 0.12	0.62 ± 0.23	12.15 ± 3.02
	z	0.52 ± 0.32	0.46 ± 0.21	26.58 ± 9.02	0.86 ± 0.11	0.75 ± 0.33	11.16 ± 3.76

Left Thigh	x	0.67 ± 0.13	0.71 ± 0.19	16.22 ± 4.39	0.88 ± 0.10	1.64 ± 0.71	9.19 ± 3.37
	y	0.61 ± 0.23	0.48 ± 0.16	23.64 ± 8.01	0.75 ± 0.23	1.12 ± 0.59	11.45 ± 4.90
	z	0.95 ± 0.05	0.40 ± 0.20	8.56 ± 4.03	0.84 ± 0.12	1.17 ± 0.59	9.79 ± 3.23
Left Shank	x	0.83 ± 0.11	0.63 ± 0.17	10.5 ± 3.67	0.96 ± 0.03	1.51 ± 0.70	5.32 ± 1.90
	y	0.85 ± 0.21	0.33 ± 0.15	11.79 ± 7.33	0.81 ± 0.19	1.51 ± 0.82	9.82 ± 4.95
	z	0.98 ± 0.02	0.47 ± 0.18	5.22 ± 1.81	0.92 ± 0.05	1.31 ± 0.51	8.00 ± 2.12
Left Foot	x	0.39 ± 0.41	1.02 ± 0.34	20.24 ± 8.14	0.95 ± 0.04	2.46 ± 1.02	5.66 ± 2.11
	y	0.98 ± 0.02	0.69 ± 0.27	4.69 ± 1.53	0.85 ± 0.15	2.15 ± 1.15	8.39 ± 4.08
	z	0.85 ± 0.17	0.79 ± 0.24	11.16 ± 4.49	0.87 ± 0.08	2.33 ± 1.10	9.30 ± 2.65
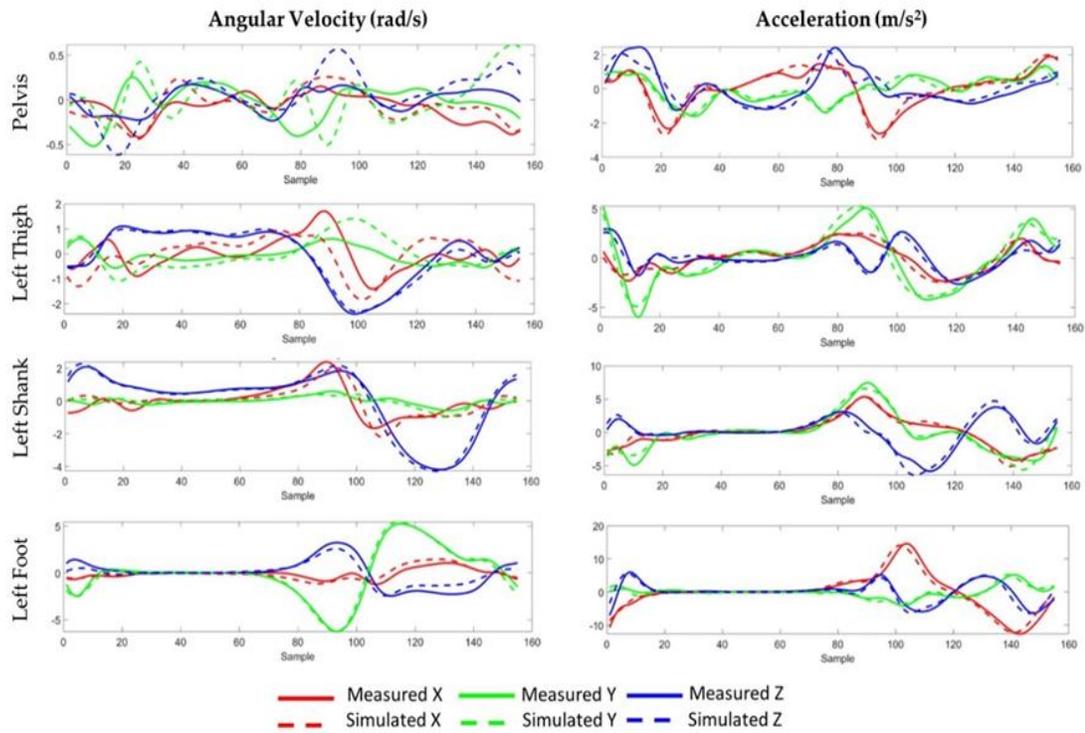
Figure 4.4: Measured (solid) and simulated (dashed) 3-D angular velocity and acceleration data during an exemplary gait cycle. The IMU sensors were located on the pelvis, the lateral left thigh, the lateral left shank, and on top of the left foot.

4.4.2 Model Accuracy

Inclusion of synthetic kinematics to supplement the measured data in the neural network training dataset statistically significantly improved kinematic predictions for all

74

hip and knee degrees of freedom (p < 0.001, Table 4.3, Figure 4.5). Likewise, the neural networks trained exclusively on synthetic data significantly improved prediction accuracy compared to models trained exclusively on measured data for five of the six kinematic degrees of freedom, excluding knee adduction–abduction (Ad–Ab) (p < 0.002). The mean RMSE and correlation coefficients for hip kinematics improved from 4.5° ± 1.6° and 0.82 ± 0.13 when trained on measured data to 2.3° ± 0.3° and 0.91 ± 0.08 when trained on synthetic data, corresponding to a 38% reduction in RMSE and a 13% increase in the correlation coefficient. Predictions improved to 1.9° ± 0.2° and 0.96 ± 0.03 when trained on both measured and synthetic data together, corresponding to a 54% reduction in RMSE and a 20% improvement in correlation coefficient, compared to the measured data alone. Mean RMSE and correlation coefficients for knee kinematic predictions followed a similar trend, improving from 3.3 ± 0.2° and 0.83 ± 0.12 when trained on experimental data to 2.9 ± 0.7° and 0.84 ± 0.12 for synthetic data, and 1.7 ± 0.4° and 0.96 ± 0.04 for the combined training dataset.

Across all joint angle predictions, the models consistently had the highest accuracy when predicting knee flexion–extension (Flex–Ext) with correlation coefficients of greater than 0.99 ± 0.01 and nRMSE ranging from 1.9 ± 0.7 to 3.9 ± 1.6 across all training datasets. Conversely, the hip and knee internal–external (Int–Ext) rotation predictions consistently had the lowest accuracy, with nRMSE ranging from 9.8 ± 3.5 to 23.9 ± 11.1 for the hip, and from 14.1 ± 6.4 to 25.5 ± 6.3 for the knee. The models trained on both measured and synthetic data had the highest generalizability with the lowest standard deviations in prediction errors for patients in the test cohort (maximum standard deviations were 0.07

for r, 1.6° for RMSE, and 6.8 for nRMSE). Accuracy metrics for individual subjects in the

test cohort are reported in Appendix B1.
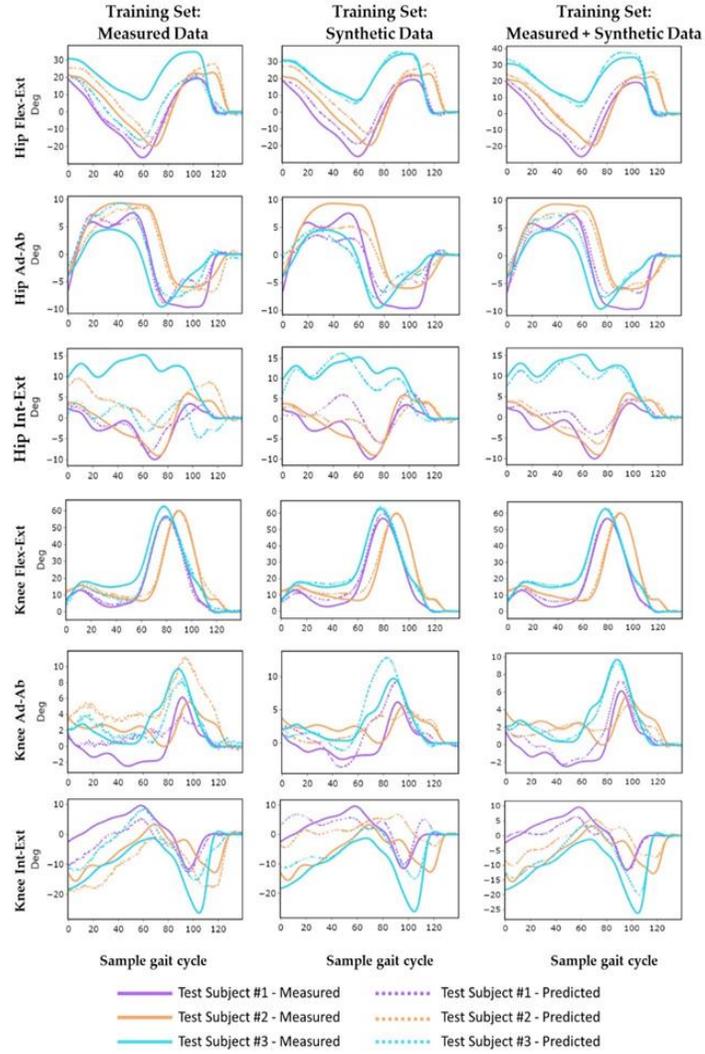
Figure 4.5: Measured and predicted lower limb kinematics in each degree of freedom during an exemplary trial for each test subject during gait: prediction (dashed line) and measured (solid line) for different training datasets.

Table 4.3: Model prediction accuracy for hip and knee joint angles with different sets of training data: measured data, synthetic data, and combined measured and synthetic data. Accuracy metrics include the Pearson correlation coefficient (r), root mean square error (RMSE), and normalized root mean square error (nRMSE).

Training Set	# Samples	Hip Flex-Ext			Hip Ad-Ab			Hip Int-Ext			Hip Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.88 ± 0.12	7.2 ± 5.0	15.4 ± 10.8	0.94 ± 0.04	2.1 ± 0.7	10.1 ± 3.2	0.64 ± 0.24	4.2 ± 2.0	23.9 ± 11.1	0.82 ± 0.13	4.5 ± 1.6	16.5 ± 8.4
Synthetic	17255	0.98 ± 0.03	2.6 ± 1.5	5.7 ± 3.2	0.95 ± 0.05	2.0 ± 0.6	9.5 ± 2.9	0.81 ± 0.17	2.3 ± 0.8	12.8 ± 4.6	0.91 ± 0.08	2.3 ± 0.3	9.3 ± 3.6
Measured + Synthetic	20706	0.98 ± 0.01	2.6 ± 1.1	5.5 ± 2.3	0.98 ± 0.02	1.3 ± 0.5	6.1 ± 2.2	0.93 ± 0.07	1.7 ± 0.6	9.8 ± 3.5	0.96 ± 0.03	1.9 ± 0.2	7.1 ± 2.7

Training Set	# Samples	Knee Flex-Ext			Knee Ad-Ab			Knee Int-Ext			Knee Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.99 ± 0.01	2.9 ± 1.1	3.9 ± 1.6	0.75 ± 0.22	2.0 ± 0.8	15.2 ± 6.3	0.77 ± 0.14	7.0 ± 1.8	25.5 ± 6.3	0.83 ± 0.12	3.3 ± 0.2	14.9 ± 4.7
Synthetic	17255	0.99 ± 0.01	2.1 ± 0.6	2.9 ± 0.8	0.82 ± 0.13	2.0 ± 0.6	15.1 ± 4.5	0.70 ± 0.24	6.4 ± 2.8	24.0 ± 11.3	0.84 ± 0.12	2.9 ± 0.7	14.0 ± 5.5
Measured + Synthetic	20706	0.99 ± 0.01	1.4 ± 0.5	1.9 ± 0.7	0.94 ± 0.06	1.2 ± 0.4	6.6 ± 2.3	0.93 ± 0.07	3.8 ± 1.6	14.1 ± 6.4	0.96 ± 0.04	1.7 ± 0.4	7.5 ± 3.1

4.5. Discussion

This study demonstrated a musculoskeletal modeling-based workflow to generate synthetic kinematic data that were used to improve the performance of neural networks to predict 3-D hip and knee rotations during gait. Supplementation of the measured kinematic training data with synthetic data that had been augmented in both magnitude and timing reduced the prediction RMSE by 54% at the hip and 45% at the knee. Training the model with synthetic data resulted in prediction accuracy that was either equivalent to or better than training purely on experimentally measured data for all three kinematic degrees of freedom at both the hip and knee.

Synthetic data used in model development must preserve the physical relationship between joint rotations and the corresponding IMU data. The musculoskeletal workflow used in the current study to generate simulated IMU data had mixed results across different limb segments. We saw the worst correlations between measurements and simulations in the pelvis rotational velocities. The magnitudes of the pelvis rotational velocities were also considerably smaller than the other limb segments during gait. Inclusion of activities that require greater pelvic rotational velocities into the training set may improve overall predictive accuracy. Additionally, rigidly attaching the pelvis IMU was challenging given the amount of soft tissue present on some study participants. These factors likely caused soft tissue artifacts to have a larger effect on the pelvis IMU measurements than the other limb segments. Angular velocity correlation coefficients for the other limb segments averaged greater than 0.79 across the three degrees of freedom. Despite the limitation associated with synthetic pelvic rotational velocities, the inclusion of the synthetic data

considerably improved the predictive ability of the model, even for the hip joint angles. This result suggests that some components of the IMU signals contribute less to the overall prediction accuracy. Future work to determine the most basic set of IMU data necessary to accurately predict joint rotations would be valuable for guiding hardware development for commercial systems used in performance monitoring or rehabilitation.

Data augmentation is commonly used to expand training datasets for machine learning algorithms. Most augmentation approaches for IMU-based applications employ label-preserving transformations such as adding noise or simulating variation in sensor positioning (Um et al. 2017). More recently, Dorschky et al. published a non-label preserving augmentation method in which planer musculoskeletal models with an optimal control simulation framework generated synthetic IMU data coupled with 2-D joint kinematics and kinetics(Dorschky et al. 2020). The optimal control simulation was necessary to preserve the physical relationship between joint kinematics, ground reaction forces, and the calculated joint kinetics but required computationally expensive and specialized modeling techniques, which may not be necessary for improving kinematic prediction accuracy. In the current study, we randomly augmented the joint kinematics with variations in both time and magnitude but did not implement controls to ensure the resulting joint kinematics were physically realistic. Additionally, the augmentation was implemented using a freely available musculoskeletal modeling framework. Therefore, these methods may be easier to implement for non-specialists in biomechanics and provide substantial time savings in the generation of valuable synthetic data for other applications. While the inclusion of unphysiological synthetic data in the training set still enabled

significant improvements of prediction accuracy, it is unclear if this would hold true of other types of movements. Future work should consider methods to generate more targeted and realistic synthetic data that span the variability in the subject population of interest such as the meta-learning algorithm proposed by Ruiz et al. (Dorschky et al. 2019; Ruiz et al. 2018).

Differences in subject characteristics, activities, dataset size, sensor type and configuration, and variation in reported data make direct comparisons of prediction accuracy with previous studies tenuous. Mundt et al. reported 3-D hip and knee joint angle predictions from a large cohort of over 88,000 cycles of simulated IMU data validated with more recent experimental IMU measurements and achieved comparable results to the current study (Table 4.4 and Table 4.5) with approximate average hip and knee RMSEs of 1.9° and 1.7°, respectively (Mundt et al. 2020b). Similar to the current study, they also observed higher prediction correlations for hip and knee flexion–extension, compared to rotations in the frontal and axial planes. While both studies used an LSTM model architecture, the current study required significantly fewer experimental observations to achieve an equivalent level of prediction accuracy. Rapp et al. predicted hip and knee joint angles during gait in 420 subjects using an LSTM model coupled with an optimization algorithm to account for differences in the predicted and measured segment rotational velocities (Rapp et al. 2021). When evaluated on simulated IMU data, the combined algorithm achieved an RMSE of 4.2° at the knee and 4.1° at the hip prior to a calibration step, which further improved the accuracy. Dorschky et al. reported RMSEs in hip and knee flexion of 5.1° and 4.8°, respectively, which were higher than the current study, but

80

included predictions for both gait and running at different speeds using a smaller training set of only 418 measured cycles and individual model for each joint angle prediction(Dorschky et al. 2020). Gholami et al. used a single IMU mounted to the foot of 10 subjects to predict hip and knee flexion during treadmill running with RMSEs of 5.6° and 6.5°, respectively (Gholami et al. 2020). It remains unclear whether the higher accuracy achieved in the current study was due to the larger experimental dataset or that higher accelerations and rotational velocities during running make predictions more difficult.

Table 4.4: Reported prediction accuracy compared with previous studies for hip kinematics. Sensor configurations included pelvis (P), thigh (T), shank (S), and foot (F). Synthetic data were data generated by augmented kinematics data, while simulated data were generated by existing motion capture data.

DoF	Reference	Sensor Configuration	# Subjects	# Cycles	Data Type	Activity	r	RMSE(°)	nRMSE
Hip Flex-Ext	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.98	2.6	5.5
	Mundt 2020a (PS-Net) [44]	P S	115	88,067	Simulated	Gait	0.98	1.6	NR
	Mundt 2020b (FFNN) [26]	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.99	5.2	NR
	Mundt 2019 (FFNN) [23]	P T S F	75	1,028	Simulated	Gait	0.99	1.3	NR
	Dorschky 2020 (CNN) [27]	P T S F	7	418 + 6688	Measured + Synthetic	Gait and Running	1	5.1	NR
	Rapp 2021 (LSTM) [43]	P T S F	420	NR	Simulated	Gait	NR	4.3	NR
	Gholami 2020 (FFNN) [10]	F	10	*NR*	Simulated	Running	0.8	5.6	9.9
Hip Ad-Ab	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.98	1.3	6.1
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.94	0.9	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.96	2.1	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.98	1.3	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NR	2.7	NR
Hip Int-Ext	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.93	1.7	9.8
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.64	2.1	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.88	5.2	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.86	2.5	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NR	5.2	NR
Hip Average	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.96	1.9	7.1
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.85	1.5	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.94	4.2	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.94	1.7	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NR	4.1	NR

Table 4.5: Reported prediction accuracy compared with previous studies for knee kinematics. Sensor configurations included pelvis (P), thigh (T), shank (S), and foot (F). Synthetic data were data generated by augmented kinematics data, while simulated data were generated by existing motion capture data.

DoF	Reference	Sensor Configuration	# Subjects	# Cycles	Data Type	Activity	r	RMSE(°)	nRMSE
Knee Flex-Ext	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.99	1.4	1.9
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.99	1.7	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.98	4.5	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.99	1.4	NR
	Dorschky 2020 CNN	P T S F	7	418 + 6688	Measured + Synthetic	Gait & Running	0.99	4.8	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NR	3.1	NR
	Gholami 2020 FFNN	F	10	*NR*	Simulated	Running	0.93	6.5	6.5
Knee Ad-Ab	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.94	1.2	6.6
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.95	1.5	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.80	2.5	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.79	1.6	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NR	3.2	NR
Knee Int-Ext	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.93	2.8	14.1
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.93	2.5	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.97	5.5	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.95	1.7	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NA	6.4	NR
Knee Average	Current	P T S F	27	3,943 + 17,255	Measured + Synthetic	Gait	0.96	1.7	7.5
	Mundt 2020a PS-Net	P S	115	88,067	Simulated	Gait	0.95	1.9	NR
	Mundt 2020b FFNN	P T S	93	3,098 + 46,437	Measured + Simulated	Gait	0.92	4.2	NR
	Mundt 2019 FFNN	P T S F	75	1,028	Simulated	Gait	0.91	1.6	NR
	Rapp 2021 (LSTM)	P T S F	420	NR	Simulated	Gait	NA	4.2	NR

We considered kinematics from the optical motion capture system as the ground truth for training and subsequent accuracy assessment; however, uncertainty in marker placement, skin artifacts, and measurement errors limit the achievable accuracy of these systems. Benoit et al. reported absolute errors in knee kinematics due to soft tissue artifacts between 2.4° and 2.8° for knee flexion, 2.5 and 4.4° for knee adduction–abduction, and 2.2° and 2.8° for knee internal–external rotations during gait (Benoit et al. 2006). In a previous study, we quantified the 5–95 percent uncertainty bounds in kinematic measurements based on input uncertainty in marker locations and movement artifacts using the current musculoskeletal modeling workflow (Myers et al. 2015). Knee flexion had the smallest uncertainty bounds ($2.7 \pm 0.3°$), while uncertainty in hip adduction–abduction ($3.0 \pm 0.3°$), internal–external ($5.1 \pm 1.0°$), and flexion–extension (6.4 ± 0.5) rotations were higher. In contrast, the RMSE in prediction accuracy achieved using the combined training set of the current study ranged from 33% to 52% of the reported uncertainty bounds for corresponding joint angles. Given this level of uncertainty, the performance of the current neural networks is well within the uncertainty in the measurement techniques. Future improvements in model performance will require higher accuracy training data.

Selection of the appropriate neural network architecture is an important step in attaining the requisite accuracy for model predictions. In previous work, we systematically evaluated multiple neural network configurations for predicting spatiotemporal gait characteristics on this same dataset and found that convolutional neural networks yielded the highest accuracy predictions (Renani et al. 2020). Mundt et al. also compared LSTM and feedforward neural networks (FFNN) performance on time-normalized gait cycle input

data and achieved better performance using the FFNN. In a more recent study, the same group compared the performance of three common neural networks for joint kinematic and kinetic predictions, finding that convolutional neural networks achieved higher accuracy than LSTM networks but required additional data processing steps that would hinder applications working in real time (Mundt et al. 2021). For the current study, we evaluated multiple network architectures (Dorschky et al. 2020; Gholami et al. 2020; Mundt et al. 2019; Zrenner et al. 2018) and found the BiLSTM model had the most robust performance. Unlike convolutional neural networks, which require input data of consistent length, recurrent neural networks such as LSTM are time independent and accept input data of arbitrary length. This approach reduced the necessary preprocessing of data (e.g., normalizing cycle to % gait) and the associated time required to build the neural networks (Dorschky et al. 2020; Mundt et al. 2020a). We used zero padding to accommodate the time dependency of the gait cycles, which led to better performance on LSTM-based networks (Hochreiter and Schmidhuber 1997; Mundt et al. 2019; Zaroug et al. 2020).

The finding that equivalent predictive accuracy can be achieved when training neural networks using synthetic kinematic data, in contrast to experimental data, expands the speed and accessibility of model development. Previously, the generation of experimental training data was the bottleneck for algorithm development and required significant investments of time and capital equipment. The current results demonstrate that reasonable predictive accuracy can be achieved using a cohort of musculoskeletal models, representative joint angles for the activities of interest, and a robust pipeline for generating simulated IMU data.

The generalizability of this workflow to additional movements, particularly more dynamic movements with higher variability (i.e., stair descent, running, or cutting maneuvers), is still unclear and requires further validation.

This study had a few notable limitations that should be considered when evaluating the results. All the subjects who participated in this study had either end-stage osteoarthritis in the hip or knee or had recently recovered from a total joint arthroplasty. This patient population has been shown to exhibit gait adaptations, including a slower pace, shorter step length, reduced knee flexion, and increased levels of variability that may affect the generalization of the model to healthy individuals (Kaufman et al. 2001; Kiss et al. 2012; Renani et al. 2020; Sparkes et al. 2019). Gait measurements were taken in the laboratory environment, which may affect the subjects' normal gait patterns. Research grade IMUs were used in this study that had on board data processing to reduce noise and drift in the signals (Vydhyanathan et al. 2015). These IMUs were placed in specific, repeatable anatomic positions to minimize variability associated with sensor positioning. Additional simulation and model training would be necessary to make the system robust to noise from lower-grade sensors and increased variability in sensor positioning on the limb segments so the system could be deployable in real-life unsupervised applications. Finally, model hyperparameter selection was based solely on the measured data and not evaluated using synthetic data. While we anticipate that incorporating synthetic data into the hyperparameter selection would further improve model accuracy when trained with synthetic data, this has yet to be demonstrated.

4.6. Conclusion

The present study demonstrated that recurrent neural network predictions of 3-D hip and knee angles during gait using IMU sensors can be significantly improved using synthetic kinematic and IMU data. On average, RMSEs in joint angle predictions were improved by 38% at the hip and 11% at the knee when models were trained on synthetic data, compared to measured data alone. When models were trained on both measured and synthetic data, RMSEs were reduced by 54% at the hip and 45% at the knee, compared to measured data alone. The musculoskeletal workflow described here enables future model development for other activities that have clinical significance without the burden of generating large quantities of gait lab data for model training.

CHAPTER 5: PRINCIPAL COMPONENT ANALYSIS OF OA AND TKA PATIENT ACROSS THE ACTIVITIES OF DAILY LIVING: GAIT, STAIR ASCENT, STAIR DESCENT, SIT TO STAND

5.1 Abstract

Biomechanical movement analysis is an important tool for clinicians and researchers to quantitatively evaluate biomechanical adaptations following disease progression, surgical intervention, or rehabilitation. A study was conducted to use principal component analysis (PCA) method to investigate mode variation between the joint kinematics of osteoarthritis (OA) and total knee arthroplasty (TKA) patients across gait, stair ascent, and strait descent. The resulted principal components (PCs) were leveraged to generate realistic synthetic kinematic data for potential training of machine learning (ML) algorithms.

Twenty-nine individuals including twelve with OA and seventeen TKA adults participated in the study. Eight kinematic waveforms related the hip, knee, and ankle joints during gait, stair ascent and stair descent tasks were utilized to perform PCA. The PCs defining at least 75% of the variation within the data were used to reconstructed mode variation waveforms and analyze the groups differences. Multivariate normal distributions were fit to the standardized PC scores for the OA and TKA cohorts during each activity and used use to generate synthetic kinematic data for machine learning applications.

Pearson correlation between the PCs of different activities as well as patients age, heights, age, and knee score were calculated and used to establish the relationship between kinematics of different activities and patient variables.

OA patients indicated statistically significant smaller stair descent pc1, gait PC2 and PC3, and stair ascent PC4 compared to TKA group. Stair descent PC1's effect was more prominent in pelvic tilt, gait PC2 in pelvic tilt and hip flexion, gait PC3 and stair ascent PC4 in hip rotation. Gait indicated the highest correlations ($abs(r) > 0.8$) with stair activities through gait PC2 and stair descent motions' PC1. No significant correlation was observed between PCs scores and any of the patients' variables. Generated synthetic kinematics data from PCs distribution were similar to original kinematics waveforms for each groups and activities.

PCA indicated reliable performance on differentiating patient groups such as knee OA and TKA as well as generating conditional synthetic kinematic data for future ML applications.

5.2 Introduction

Osteoarthritis (OA) is one of the most common joint degenerative conditions, specifically in elderly patients, and results in movement impairment that can be observed by gait examination (Kiss et al. 2012). People with knee OA usually suffer from pain, poor quality of life, significant quadriceps weakness, and reduced functional mobility (Whitchelo et al. 2013). Total knee arthroplasty (TKA) is widely used to treat knee OA. It typically reduces knee stiffness, relieves pain, improves lower extremity alignment, and quality of life (Kiss et al. 2012; Tazawa et al. 2014). Despite the success rate of TKA, about

one in five TKA patients are not happy with their outcomes (Ardestani et al. 2017) and experience functional limitation after surgery (Whitchelo et al. 2013). These functional limitations, as well as dissatisfaction, are more apparent in high demand activities like ambulating stairs (Du et al. 2014).

Biomechanical movement analysis is an important tool for clinicians and researchers to quantitatively evaluate biomechanical adaptations following disease progression, surgical intervention, or rehabilitation (Zeni and Higginson 2009). Many of these analyses focus on level gait while a more challenging task, stair ambulation, has received less attention among people with functional impairments. Stair ascent and descent are more demanding on lower extremity muscles and joints, requiring greater ranges of motion and joint moments, resulting in an elevation of pain (Andriacchi et al. 1980). Hence, it's common to adopt compensatory strategies to alleviate pain among individuals with impaired joints, which then affects the mechanics of the other joints in the kinematic chain. Biomechanical deficiencies in stair motion can make apparent factors which are important for understanding movement adaptations OA and TKA patients (Meyer et al. 2016; Standifird 2015).

Despite a wealth of information collected from movement analysis, interpretation of these analyses is complicated due to the interdependency of biomechanical movement characteristics. The conventional method for interpreting kinematic and kinetic data is to focus on isolated features of the waveforms, such as peak or rate, and then statistically compare them as independent variables. The selection of these metrics is subjective and frequently varies between studies which produces inconsistent results and makes

comparison between studies difficult (Meyer et al. 2016; Standifird 2015). Further, kinematic variables are not independent and considering the interdependency is necessary for understanding movement adaptations. In contrast, principal component analysis (PCA) is a mathematical technique to reduce the dimensionality of complex data and to extract meaningful features (Cushion et al. 2019; Deluzio et al. 1997). This enables objective analysis of movement characteristics across patient populations and different activities.

The use of PCA for human movement analysis is commonplace (Chau 2001; Deluzio et al. 1997, 1999) and has been employed to detect kinematic differences between healthy and OA cohorts during gait (Deluzio and Astephen 2007; Federolf et al. 2013; Robbins et al. 2013), to associate patients' kinematics with clinical observations (Bensalma et al. 2019), and to quantify the effect of physiotherapy on knee kinematics during gait (Brenneman and Maly 2018; Gaudreault et al. 2011; Yocum et al. 2021). Biggs et al. combined PCA with classification methods to identify changes in gait characteristics associated with TKA. The analysis identified a combination of 18 gait characteristics that differentiated OA from healthy gait, and a subset of those characteristics that changed after TKA, including ground reaction forces and hip adduction moments (Biggs et al. 2019b). In subsequent analyses, variation in gait characteristics were strongly correlated with patient outcomes like the Oxford Knee Score (Biggs et al. 2019b).

Due to a surge of data and flourishing advances in machine learning (ML), recent efforts have employed ML for biomechanical analysis. ML can model underlying relationships between movement strategies with broad applications from predicting disease progression (Brenneman and Maly 2018; Gaudreault et al. 2011; Yocum et al. 2021) to

optimizing clinical interventions (Biggs et al. 2019b, 2019a). The primary barrier to employing ML in biomechanics is the lack of ample high quality data due to the difficulties of measuring patient movement and patient privacy concerns (Myers et al. 2018). Generation of realistic synthetic movement data could help to alleviate this challenge. A limited number of studies have used synthetic IMU data to train ML models that monitor gait and have shown significant improvements in the models' predictive capabilities (Dorschky et al. 2020; Mundt et al. 2020b; Renani et al. 2021). Developing a comprehensive process for synthetic data generation remains an unmet need in the field. In addition to the ability to identify complex movement differences between patient cohorts, PCA could potentially be leveraged as an autoencoder for realistic synthetic data generation that preserves the unique compensatory strategies employed by OA and TKA patients (Trinler et al. 2016).

To our knowledge, PCA has not been used to analyze higher demand activities like stair ascent and descent or to investigate movement strategies that exist across stair ambulation and gait in OA and TKA patients. Further, PCA has not previously been used to generate realistic synthetic kinematic data for potential training of ML algorithms. Therefore, the purpose of the current study are two-fold: first, to identify the modes of variation for OA and TKA in during stair ascent, stair descent, and gait, and second, to use PCA to provide an open-source tool for generating synthetic kinematic data for future ML applications and musculoskeletal analyses.

5.3 Methods

5.3.1 Experimental Data Collection

Twentynine subjects participated in this study, including 12 subjects with OA (age = 63 ± 6, weight = 76 ± 14 kg, height = 165 ± 13 cm, 6 females and 8 males) and 17 subjects with TKA (age = 68 ± 5, weight = 76 ± 14 kg, height = 163 ± 13 cm, 13 females and 4 males). Patients in the OA cohort had diagnoses confirmed by orthopaedic surgeons via radiographic review and were deemed candidates for TKA. Patients in the TKA cohort were at least one-year post TKA surgery. Prior to the experiment, participants were asked to complete the Knee Outcome and Osteoarthritis Score (KOOS) and sign a consent form under an approved study by our Institutional Review Board (#1328728-1).

Subjects were fitted with 71 reflective markers on anatomical landmarks. Each subject performed 15 trials of a 5-meter walking task barefoot at three different speeds: self-selected, slow, and fast. During slow walking, subjects were instructed to walk at their slowest comfortable speed by taking shorter steps and slowing their step cadence. Conversely, during fast walking, subjects were instructed to walk at their maximum comfortable speed, without running, by taking longer steps at a faster pace. Subjects were also asked to ascend and descend seven steps on a purpose-built wooden staircase (step height × length × width = 7" × 5.75" × 22.25") six times in each direction. During stair ascent and descent trials, subjects were instructed to place one foot on each step using a self-selected pace and were allowed to grasp the handrail if necessary to aid balance. A Vicon motion capture system (Centennial, CO) with thirteen cameras (Columbus, OH)

were used to collect three-dimensional marker positions attached to each body segment at the rate of 100 Hz during all activities.

5.3.2 Musculoskeletal Modeling

Using a previously published workflow, subject-specific musculoskeletal models were generated for each participant. Each model included 22 rigid body segments with 26 joints. The pelvis and hip joints were modeled with 3 rotational degrees of freedom each, and the knees and ankles as single degree of freedom hinge joints. Limb segments were scaled to match the optical markers from the experiment. An inverse kinematics analysis of each trial was performed in OpenSim to obtain kinematics of the pelvis, hip, knee, and ankle. The procedure was conducted by solving a weighted least squares optimization problem with the goal of minimizing the distance between an experimental marker and the corresponding virtual marker. The weight values of 1 was selected across all markers. Joint angles calculated from the measured data were segmented into individual cycles using the heel marker, resulting in a total of 2652 unique cycles (gait: 1843, stair ascent: 351, and stair descent: 353). The segmented kinematic data were normalized to 100 samples and the mean kinematic profiles for each subject performing each activity across trials were calculated.

5.3.3 Principal Component and Statistical Analyses

Separate PCAs were performed on the assembled patient mean kinematic data for gait, stair ascent, and stair descent (Myers et al. 2018). The principal components (PCs) defining at least 75% of the variation within the data were selected for further analysis. Representative modes of variation were reconstructed by adding ±2 standard deviations of

each PC score (e.g. eigenvalue) to the mean score, multiplying by the corresponding PC loading vector (e.g. eigenvector), then adding to the mean kinematic profiles (Brandon et al. 2013). The reconstructed kinematic waveforms of each PC were compared with the mean kinematics to highlight the features of each mode of variation. Independent samples t-tests were used to detect significant differences in PC scores between the OA and TKA cohorts and between female and male cohorts ($\alpha=0.05$) (Brandon et al. 2013). Absolute value of Pearson correlations coefficients were calculated between each subject's PC scores and their weight, height, and KOOS scores. Similarly, Pearson correlation coefficients were calculated between each subjects PC scores in gait, stair descent, and stair ascent.

5.3.4 Statistical Analysis - Group Difference

To identify characteristic group differences between the patterns of the OA and TKA patients as well as female and male subjects, independent statistical analysis of t-tests ($\alpha=0.05$) was employed on Ci scores across different activities (Reid et al. 2010). We further examined PCs with significant effects by examining corresponding mode variations and LVs reconstructed in the previous section. We also investigated the relationship between PCs and other patient demographic variables by calculating absolute Pearson correlations between PC, Ci, and age, weight, height, and KOOS scores. In a similar manner, to determine their relationship between kinematic profiles, absolute Pearson correlation coefficients between the Ci scores of different activities were calculated and displayed using heatmaps.

5.3.4 Synthetic Kinematic Data Generation

To generate synthetic kinematics data from the existing PC analyses, multivariate normal distributions were fit to the standardized PC scores for the OA and TKA cohorts during each activity. The means and covariance matrices of these normal distributions were calculated. New standardized PC scores were then sampled from these distributions and transformed into the original scale by multiplying with the PC score's standard deviation and adding to the PC score's mean. New synthetic kinematics were then generated by multiplying the sampled PC scores with the PC loading vectors and adding to the mean kinematic profiles.

5.4 Results

5.4.1 Modes of variation during gait, stair ascent, and stair descent

During gait, the first four PCs accounted for 78% of the variance in kinematics. PC1 (31% of variation) described increased pelvis list, hip abduction, and knee and ankle flexion during stance. PC2 (23% of variation) described increased pelvic tilt and hip extension throughout the gait cycle and pelvic list, hip adduction, and knee extension during swing. PC3 (15% of variation) described increased hip external rotation throughout gait and peak hip abduction during swing. Finally, PC4 (9% of variation) described pelvis tilt, pelvis rotation, and hip abduction during stance and increased knee flexion in swing.

During stair ascent, the first five PCs accounted for 77% of the variance in kinematics. PC1 (23% of variation) primarily described increased forward pelvic tilt and hip flexion throughout the cycle. PC2 (18% of variation) described increased hip abduction during stance, increased knee flexion at toe-off, and pelvic rotation throughout the cycle, coupled

with. PC3 (17% of variation) described increased pelvic rotation at heel strike, on and hip rotation, end of the cycle of hip rotation, and at the middle phase of ankle angle. PC4 (12%) was mainly located at the end cycle of hip adduction, middle of hip rotation, and beginning and end of knee flexion. PC5 (8%) was located at hip abduction, hip rotation, and knee angle (Figure C1. b).

Similarly, five components were selected to cover above 75% variation within the dataset in the stair descent activity. PC1 (26%) was mostly located in the pelvic tilt during the entire stance phase, in the pelvis list during the end of stance phase and beginning of the swing phase, and in the ankle angle during beginning of stance phase and end of swing phase. It was also observed in the hip flexion and adduction during control lowering/double support of stance phase, weight acceptance and forward continuance, and end of controlled lowering subphase of stance phase, as well as foot placement of swing phase. PC2 (20%) was noticeable at the time of foot placement during the end of the swing phase in the pelvis list and the entire phase of the pelvis and hip rotation. PC3 (17%) was responsible for the highest variation in pelvis tilt and hip flexion of stance phase, specifically during weight acceptance and beginning of single support phase within pelvis list and hip flexion. It was also observed during the swing phase of hip and knee flexion and partially in ankle dorsiflexion. PC4 (10%) was mainly seen in knee flexion during the entire stance phase. PC5 (7%) was seen in the pelvis list during the beginning of the stance phase, hip abduction during the beginning of the stance phase, and foot placement of the swing phase (Figure C1.C).

5.4.3 PCs Component Group Difference

When compared with TKA, OA patients had lower PC1 scores in stair descent, lower PC2 and PC3 scores in gait activity. As a result, the reconstructed waveforms that represent the low score percentiles, - 2 standard deviations, were representative of PC1, and PC2, and PC3 effects on OA patients in gait and stair activities, whereas the high scores were representative of PC effects on TKA patients. Similarly, females had significantly lower PC4 scores in gait, PC5 scores in stair descent, and larger PC4 scores in stair ascent compared to males (Table. 5.1).

Table 5.1: The PCs values across different activities for knee groups (OA and TKA) and gender (male and female)

PC variables	TKA	OA	Male	Female	Knee p-value	Gender p-value
Gait_pc1	-31.4 ± 92.9	38.6 ± 90.3	35.6 ± 120.8	-21.7 ± 74.5	0.051*	0.124
Gait_pc2	31.3 ± 76	-38.5 ± 80.2	-1.3 ± 81	0.8 ± 88.5	0.023**	0.949
Gait_pc3	24.6 ± 66.3	-30.3 ± 57	-28 ± 57.6	17.1 ± 68.4	0.026**	0.080
Gait_pc4	-8.9 ± 42.4	11 ± 61.8	27.7 ± 64	-17 ± 35.2	0.314	0.022**
StairAscent_pc1	-26.7 ± 84.4	32.9 ± 103.1	1.4 ± 65.3	-0.8 ± 113	0.098	0.954
StairAscent_pc2	-27.6 ± 54	34 ± 105.4	36 ± 112.5	-22 ± 56.7	0.052*	0.075
StairAscent_pc3	-7.5 ± 84.6	9.2 ± 80.8	-4.9 ± 94.7	3 ± 75.8	0.593	0.808
StairAscent_pc4	23.2 ± 63.5	-28.6 ± 67.3	-34.4 ± 76.9	21.1 ± 56.3	0.043**	0.033**
StairAscent_pc5	-2.7 ± 59.4	3.4 ± 58.1	25.8 ± 69	-15.7 ± 44.9	0.784	0.059*
StairDescent_pc1	57.8 ± 83.3	-71.1 ± 79.3	-35.6 ± 97.6	21.8 ± 103.2	0.000**	0.150
StairDescent_pc2	7.2 ± 91.1	-8.9 ± 97.1	-40.6 ± 64.1	24.8 ± 99.8	0.649	0.064
StairDescent_pc3	21.4 ± 84.3	-26.3 ± 83.6	-38.3 ± 91.3	23.4 ± 75.7	0.139	0.059*
StairDescent_pc4	4.8 ± 46.7	-5.9 ± 82.8	-2 ± 84.5	1.2 ± 51.1	0.666	0.899
StairDescent_pc5	-4.7 ± 46.9	5.8 ± 62.3	26.9 ± 58.4	-16.4 ± 44.4	0.607	0.032**

```
** p-value ≤ 0.05
* p-value ≤ 0.06
```

5.4.3.1 PC1 stair descent

The reconstructed waveform from PC1 as well as the mean kinematic profile of OA and TKA groups were displayed in Figures 5.1. a and b. Overall PC1 was responsible for the most variance of the entire pelvic tilt, hip flexion, hip adduction, and ankle dorsiflexion.

The PC1 LVs' descriptions related to each joint were listed in Table. C3. This PC LV captured a difference feature in pelvis rotation and knee angle and magnitude feature in other joint angles. OA and TKA group kinematic profiles agreed with waveforms reconstructed by PC1 in all joint angles. OA subjects indicated higher pelvic posterior tile, more hip abduction, smaller hip flexion, smaller hip rotation, and smaller knee and ankle ROM compared to TKA.
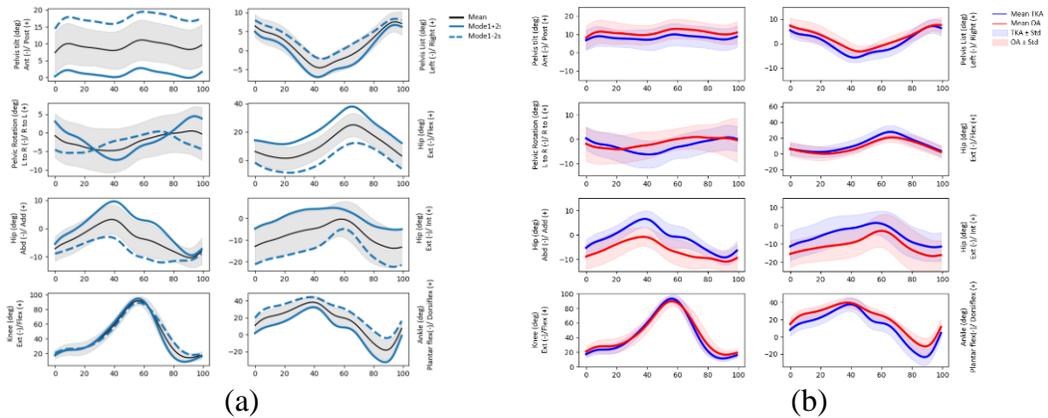
Figure 5.1: a) Representative extreme or mode of variations for each joint described by reconstructing the waveform from PC1 and by perturbing the mean kinematic profile by ± 2 standard deviations. b) the mean OA and TKA kinematic for stair descent activity.

5.4.3.2 PC2 and PC3 Gait

PC2 was more dominant in the pelvis, hip flexion, and hip adduction, while the PC3 was responsible for the most variation in hip rotation. The PC2 LV was capturing the magnitude feature in Pelvis tilt, pelvis rotation, hip flexion, and ankle angle, and the difference feature in pelvis list, hip adduction, and knee angle (Table C4). The PC3 LV is capturing the magnitude difference in the pelvis and hip rotation and difference feature in pelvis Tilt, hip adduction, and slightly in knee angle (Table C4). As shown in figure 5.2. a & b, OA and TKA group kinematic profiles agreed with waveform reconstructed by PC2

in pelvis title, pelvis list, hip adduction, and ankle angle. While they agreed with PC3 mainly in hip rotation, and some knee angle. Considering only PC2, the OA patients (low scoring PC2) indicated more posterior pelvic tilt and approximately the same hip flexion with smaller hip flexion, lower hip adduction at swing phase, and more ankle dorsiflexion. Regarding PC3, the OA patient (low scoring PC3), indicated less hip internal rotation throughout the gait cycle, lower hip adduction at swing phase, and higher peak knee flexion at swing phase.

(a)   (b)
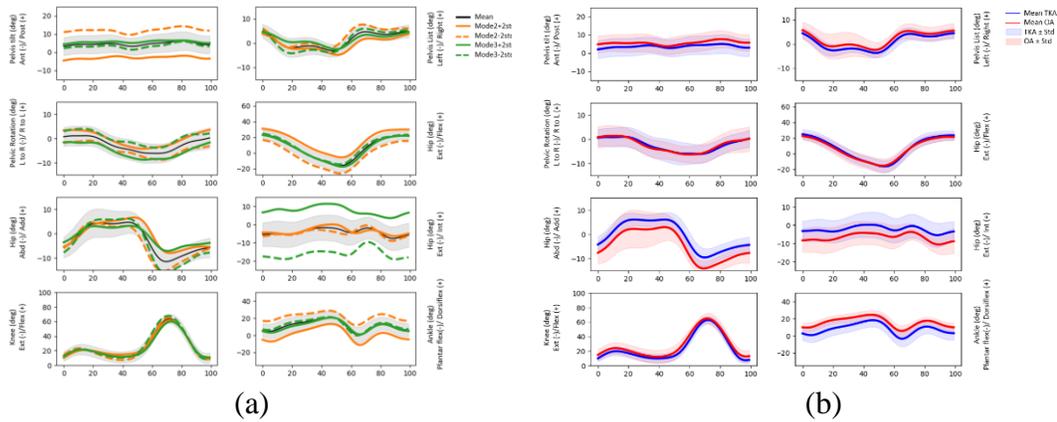
Figure 5.2: a) Representative extreme or mode of variations for each joint described by reconstructing the waveform from PC1 and by perturbing the mean kinematic profile by ± 2 standard deviations. b) the mean OA and TKA kinematic for stair descent activity.

5.4.3.3 PC4 stair ascent

In stair ascent, the PC4 was more dominant at the hip abduction, hip rotation, and knee flexion. The PC2 LV was capturing the Magnitude feature in Pelvis tilt, pelvis rotation, hip rotation, and ankle angle, and the difference feature in the other joints (Table C5). Comparison between reconstructed waveform and knee groups' profiles reveals that OA and TKA group kinematic profiles agreed with waveform reconstructed by PC4 mainly at hip rotation, knee angle, and ankle angle (Figure 5.3 a & b). In general, OA subjects also

indicated lower hip abduction and lower hip internal rotation, slightly higher knee flexion

during the stance phase, and higher ankle dorsiflexion compared to TKA.

| (a) | (b) |
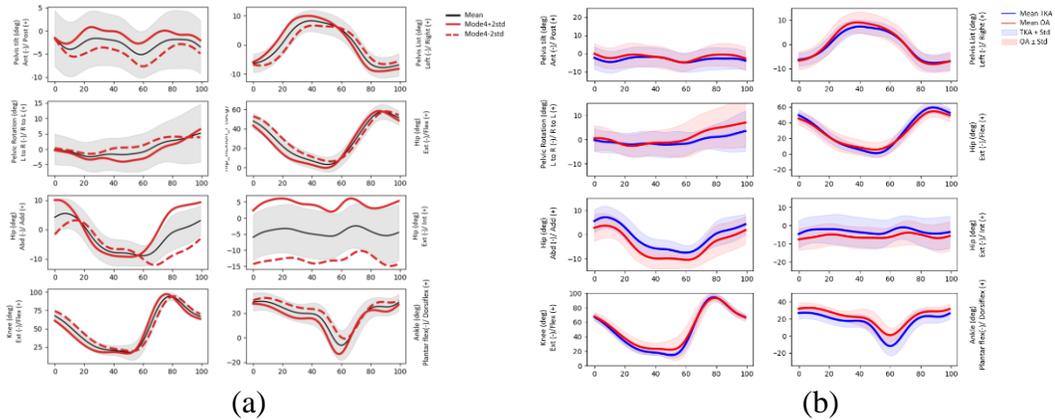
Figure 5.3: a) Representative extreme or mode of variations for each joint described by reconstructing the waveform from PC1 and by perturbing the mean kinematic profile by ± 2 standard deviations. b) the mean OA and TKA kinematic for stair descent activity.

5.4.4 PCs Correlation between PCs and Anthropometrics

Absolute correlation between PCs and subjects' anthropometric variables as well as

PCs of inter and intra-activities calculated using absolute Pearson values were shown in

figure 5.4.a. and figure 5.4.b. respectively. There wasn't any significant correlation

between patient anthropometric variables and PC variables. The highest correlation was

related to KOOS score and stair ascent PC2 and stair descent PC1. Regarding intra-

activities correlations and considering ±0.6 correlation as a threshold value, four clusters

were formed. These clusters are as follows: first, Gait pc2, Stair Descent PC 1, and Stair

Ascent PC1. Second, Gait PC4, Stair Descent PC5, and Stair Ascent PC5. Third, Gait PC1,

Stair Descent PC3, and Stair Ascent PC3. Forth, Gait PC3, Stair Descent PC2, and Stair

Ascent PC4.
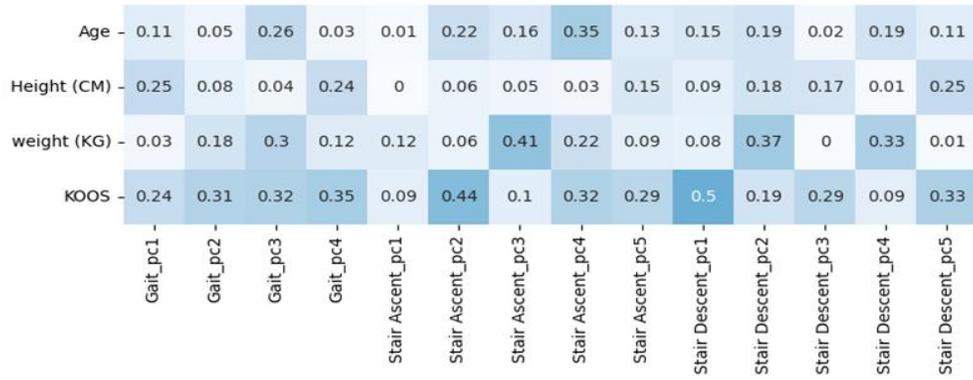
(a)
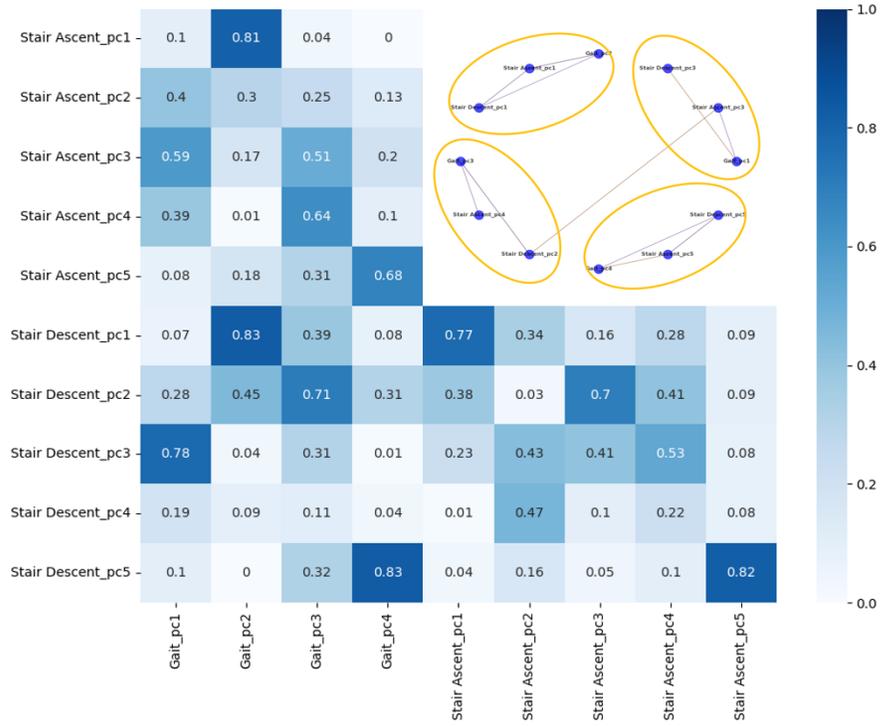
(b)

Figure 5.4: Absolute Pearson correlation displayed using a heat map matrix across a) PCs activities and patient demographic variables, and b) between PCs activities.

5.5 Discussion

The majority of previous studies tried to analyze OA and TKA kinematics, in gait or considering only knee kinematics, or they were limited to specific parameters of movement waveform such as peak values, magnitude at a specific gait cycle event, etc. The most discriminating activity between knee OA and TKA was stair descent, followed by gait and stair ascent. Similarly, Schutz et al. 2019 also indicated stair descent as the most discriminating activity between hip OA and the healthy group at a stair height of 7.2 in compared to ours at 7 in (Reid et al. 2010). Despite this result, Ursula K Trinler et al. 2018 investigated the impact of stair height on the kinematics and kinetics of TKA and control groups during stair ascent and descent using 6.7 in. and 8.3 in. stairs (Schütz et al. 2019). They revealed the significant impact of the stair configuration on kinematics and kinetics, and they emphasized the importance of considering stair height when comparing groups and studies.

Regarding gait, PC2 and PC3 were the top significant discriminators, with lower PCs scores for OA compared to the TKA group. PC2 was more dominant in the pelvis, hip flexion, and hip adduction, while PC3 was responsible for the most variation in hip rotation. According to PC2, the OA patients (low scoring PC2) indicated more posterior pelvic tilt, and approximately equal hip flexion, lower hip adduction at the swing phase, larger peak knee flexion at the swing phase, and slightly more ankle dorsiflexion. Regarding PC3, the OA patient (low scoring PC3), indicated less hip internal rotation throughout the gait cycle, lower hip adduction at the swing phase, and lower knee flexion at the stance phase. The PCs calculated in this study agreed with PCs values reported by Biggs et al. 2019 with a

larger PC1 score at hip flexion/extention, and a smaller PC2 at hip adduction/abduction, hip internal/external rotation, and knee flexion for the OA group compared to TKA (Trinler et al. 2016). They used PCA and the Cardiff classifier to define 18 biomechanical features that distinguished OA from healthy gait and whether these features were affected by TKA (Biggs et al. 2019b; Sparkes et al. 2019). Several other studies compared the gait kinematics of patients pre- and post-TKA (Biggs et al. 2019b) for different demographics and severity of knee pre-TKA (Hatfield et al. 2011; Levinger et al. 2011; Rao et al. 2022; Ro et al. 2020), total and unicondylar KA (Young-Shand et al. 2020), and gender (Agarwal et al. 2019). One common trend observed in their results was increasing knee ROM after TKA, and our results supported their finding. They also reported an increased peak knee flexion during the swing phase after TKA while we noticed similar peak flexion between the two groups. This could happen due to using two different subject groups OA and TKA instead of considering one subject group pre- and post-TKA. Hip flexion and adduction ROM and ankle dorsiflexion also followed the trend reported in Pazit Levinger 2013 and Du Hyun Ro 2020, both of which showed greater ROM with TKA compared to OA (Levinger et al. 2013; Ro et al. 2019; Wilson et al. 2015). Both OA and TKA patients followed different recovery strategies to avoid pain, maintain their balance, and perform their daily activities. These differences might be due to muscle weakness or joint stiffness around the knee caused by OA or surgery. For instance, the greater ROM in the knee among the TKA group might be attributed to the reduction in pain after the survey. An increased ankle dorsiflexion in OA patients is thought to compensate for impaired knee function and to generate sufficient power for propulsion. This compensatory response may be due to the

lengthening of the calf muscles prior to push-off, contributing to force generation during push-off. Additionally, the increased ankle dorsiflexion may also compensate for the lack of knee extension observed in the surgical group during late stance, assisting in advancing the leg into swing and propelling the body forward (Elkarif et al. 2020; Levinger et al. 2013). Unlike gait activity, there was significantly fewer data associated with the OA and TKA population performing more high-demand activities such as stair ascent or stair descent.

In stair descent, the OA group resulted in a lower PC1 score compared to TKA (p-value = 0.000). The representative extreme reconstructed from PC1 captured the OA and TKA behavior at all the joints with the most dominance at pelvis tilt, hip flexion, hip adduction, and ankle angle. Overall, OA subjects indicated larger posterior pelvic tilt, more hip abduction, smaller hip flexion, smaller hip rotation, and smaller ankle ROM compared to TKA. Regarding the knee joint, the difference between OA and TKA was not significant, but the OA group indicated a smaller peak flexion during swing time, a smaller knee extension, and a smaller ROM. Among these results, the peak knee flexion angle was in agreement with the results reported by Sumner et al. 2019(Levinger et al. 2011; Ro et al. 2020) but in contradiction with the study by Komaris et al 2021 (Sumner et al. 2019), where they found a higher peak value in the preop (OA) group compared to the post-op (TKA). The other two knee parameters were also comparable with their report, but they depended on the knee implant type. They reported that the OA group had lower and higher ROM and minimum knee angle (knee extension) compared to the TKA groups who received multi-radius vs single-radius knee implants. Similarly, Komaris et al 2021 indicated smaller

ROM in the TKA group with high congruency fixed and high congruency mobile implants, but larger ROM in TKA with low congruency fixed relative to the age-matched OA group in stair descent activity (Komaris et al. 2021).

Regarding the stair ascent, the OA group showed a significantly lower PC4 score compared to TKA (p-value = 0.043). It was more dominant in hip adduction at the swing phase and in hip rotation at the stance phase and knee joint. The representative extreme waveforms reconstructed from PC4 were visible at hip rotation, knee, and ankle angles. OA subjects also indicated lower hip abduction and lower hip internal rotation, slightly higher knee flexion during the stance phase, and higher ankle dorsiflexion. The knee ROM trend agreed with Komaris et al. 2021's results for TKA subjects with fixed-bearing implants and disagreed with subjects with mobile-bearing. Identifying the knee implant type was not in the scope of our study while recruiting subjects (Komaris et al. 2021). Observed kinematic alterations at the ankle and hip joints may be compensatory responses to allow forward momentum and sufficient power production for propulsion post-TKA (Komaris et al. 2021).

Patient demographic variables indicated various correlations with PCs of different activities. However, the PCs produced more noticeable differences in discriminating knee groups compared to gender. The knee group difference was detectable in all three activities with the largest variation at stair descent PC1 (p-value=0.000) and smallest at stair ascent PC4 (p-value=0.043). Unlike the knee, the discriminator power of the gender group's difference was smaller with a significant p-value ranging from 0.022 for gait PC4 to 0.033 for stair ascent PC4. In regard to anthropometric variables, age, height, and weight didn't

have any high correlation (r>0.5) with PCs. On average, the weight indicated the highest number of larger correlation values with PCs activities than age and heights respectively. The weight' high correlation values were related to gait pc3, stair ascent PC3, and stair descent PC2 and PC4. Previous studies analyzed the effect of speed, gender, age, and BMI on gait kinematic and kinetic data (Levinger et al. 2013). They reported an increased hip flexion and anterior pelvic tilt during gait among females (Chehab et al. 2017; Moissenet et al. 2019). They also reported the significant effects of BMI on sagittal hip kinematics with an increased hip flexion as BMI increased. Rosso et al. 2019 (Asai et al. 2017) reported a high correlation between BMI in the frontal and transverse planes for normal and obese groups. In our study, PC3 gait, the highest correlated PC with weight, was more dominant in pelvis and hip rotation as well. Height did not have any high correlation with any of the PCs. This allows synthetic kinematic generation based on solely patient knee and gender. KOOS score had the highest correlation with PCs of activities, specifically with stair descent PC1. KOOS and knee status (OA=0 and TKA=1) also indicated a high and negative correlation to each other. They were also mostly correlated to top discriminator PCs, indicating their main responsibility in driving the variability of the kinematic waveform and associated PCs. In other words, lower KOOS score associated with knee OA correlated with changes in waveform profile reconstructed by gait PC2 and PC3, stair descent PC1, and stair ascent PC2.

The PCs calculated from each knee group within each activity were used in fitting two Gaussian distributions, OA and TKA, which were then used to draw samples for generating new kinematic data via an autoencoder PCA-based framework. As a result, six distributions

were generated, where PC sampled from each distribution were then used to generate kinematic data about the activity and knee group. The correlation's results of PCs and patients' demographic and anthropometric variables revealed that age, weight, and height variables may not be as critical as knee and gender for generating synthetic data. This finding supports Moissent et 2019. 's study where they used a multi-linear regression model to predict lower limb kinematics from speed, gender, age, and BMI (Rosso et al. 2019). They reported speed as the most significant factor followed by gender, BMI, and lastly age in generating synthetic data. In this work, because of a low sample of data points, we only focused on generating synthetic kinematics based on knee status using linear autoencoder. For future studies, we recommend enlarging the data set and incorporating gender status as additional conditional factors. Additionally, other nonlinear autoencoder techniques should be explored for generating conditional synthetic data. Some of these techniques include neural network autoencoders, variational autoencoders, and generative adversarial networks (Moissenet et al. 2019).

The OA and TKA groups indicated a relatively consistent trend across five out of eight joints among the activities with different discriminator power. These consistent trends for the OA group included more pelvis posterior tilt, pelvis list toward standing leg, less hip adduction, hip rotation, and more ankle dorsiflexion. The three other joints with inconsistent trends were pelvis internal rotation, hip flexion, and knee angle. The kinematics of the OA and TKA groups were similar at pelvis rotation and hip flexion in gait activity while they were different in stair motion activities. Particularly, OA groups produced more internal pelvis rotation and less hip flexion, smaller knee ROM, knee

extension, and peak knee flexion in stair activities compared to TKA. This perpetuated trend between activities could be a movement strategy to avoid pain and compensate for the functional impairment by engaging the pelvis as a primarily driven joint instead of the knee among OA groups when performing stair motion. Christophe A.G. Meyer also reported that most stair motion features associated with hip OA were similar to gait. They showed prominence of a decreased hip flexion and increased trunk later flexion toward the affected side among hip OA in stair motion activity (Bond-Taylor et al. 2021; Das et al. 2021).

Different levels of correlation were noticed across different PCs of activities. Gait indicated the highest correlations (abs(r) > 0.8) with stair activities through gait PC2 and stair motions' PC1. These correlations were signified in pelvis tilt and hip flexion. Similar correlations were observed between stair descent and ascent PC1s with high correlation highlighted at pelvis tilt, hip flexion, and ankle dorsiflexion. Stair ascent also indicated a high correlation with stair descent and gait at hip adduction and knee respectively via stair motion PC5 and gait PC4.

The study had several limitations, including the fact that the sample size (n = 29) is small when compared with the number of statistical inferences made, which may increase the chance of an error. A further limitation was the different subjects associated with the OA and TKA groups. In previous studies, they analyzed the biomechanic changes of OA patients before and after TKA. Although we tried to control the effect of patient demographic variables on our results by collecting a relatively similar population, there were some differences between the two groups, such as age and weight. Additionally,

subjects' implant designs have not been considered for this study, which may affect the kinematic profile. The implant designs indicated alternation of the gait biomechanics as described in the literature (Meyer et al. 2016).

5.6 Conclusions

In conclusion, PCA indicated promising performance in differentiating patient knee groups as well as a new approach to generating conditional synthetic kinematic data. Our finding provided clear evidence that OA and TKA patients can be more distinguishable when they were being analyzed in more high-demand activities, such as stair descent. Our results demonstrated that considering the whole lower extremity joint or even the whole body in the analysis can produce more reliable outcomes, especially when we are dealing with low sample data.

CHAPTER 6: BioMAT: AN OPENSOURCE BIOMECHANICS MULTI-ACTIVITIES TRANSFORMER FOR JOINT KINEMATIC PREDICTION BASED ON IMU

6.1 Abstract

Through wearable sensors and deep learning techniques, biomechanical analysis can reach beyond the lab for clinical and sporting applications. Transformers, a recent deep learning model, have become widely used in state-of-the-art artificial intelligence research due to their superior performance in various natural language processing and computer vision tasks. The performance of transformer models has not yet been investigated in biomechanics applications. In this study, we introduce a biomechanical multi-activity transformer-based model, BioMAT, for the estimation of joint kinematics from streaming signals of multiple inertia measurement units (IMUs) using a publicly available dataset. This dataset includes IMU signals and the corresponding sagittal plane kinematics of the hip, knee, and ankle joints during multiple activities of daily living. We evaluated the model's performance and generalizability and compared it against a convolutional neural network long short-term model, a bidirectional long short-term model, and multi-linear regression across different ambulation tasks including level ground walking (LW), ramp ascent (RA), ramp descent (RD), stair ascent (SA), and stair descent (SD).

To investigate the effect of different activity datasets on prediction accuracy, we compared the performance of a universal model trained on all activities against task-specific models trained on individual tasks. When the models were tested on three unseen subjects' data, BioMAT outperformed the benchmark models with an average root mean square error (RMSE) of $5.5 \pm 0.5°$, and normalized RMSE of $6.8 \pm 0.3°$ across all three joints and all activities. A unified BioMAT model demonstrated superior performance compared to individual task-specific models across four of five activities. The RMSE values from the universal model for LW, RA, RD, SA, and SD activities were $5.0 \pm 1.5°$, $6.2 \pm 1.1°$, $5.8 \pm 1.1°$, $5.3 \pm 1.6°$, and $5.2 \pm 0.7°$ while these values for task-specific models were, $5.3 \pm 2.1°$, $6.7 \pm 2.0°$, $6.9 \pm 2.2°$, $4.9 \pm 1.4°$, and $5.6 \pm 1.3°$, respectively. Overall, the proposed BioMAT model accurately estimated joint kinematics across different activities directly from the sequence of IMUs signals instead of time-normalized gait cycle data.

6.2 Introduction

Accurate measurement and prediction of joint kinematics enable the development of tools for pathological diagnosis, implant design, rehabilitation, sports science, and ergonomics (DORR et al. 1988; Hantouly et al. 2022; Ishii et al. 1998; Migliorini et al. 2022; RITTMAN et al. 1981). Passive-marker motion capture (MOCAP) systems are the current gold standard in measuring joint kinematics. However, the use of these systems is time-consuming, restricted to lab environments, and requires technical expertise (Al-Zahrani and Bakheit 2002; Baker 2006; Ryan et al. 2021; Sartori et al. 2016b). In contrast, wearable inertial measurement units (IMUs) have gained attention in biomechanics

applications and joint kinematic measurement due to their portability, ease of use, and low cost.

Deep learning, a subset of machine learning, has significantly advanced the capability in converting IMU signals into joint kinematics. Among those, Mundt et al. tested various deep neural network (NN) models, including multi-layer perceptron, convolutional neural network (CNN), and recurrent neural networks (RNNs) such as long short-term memory (LSTM) models, in their ability to estimate joint kinematics and kinetics from measured IMU signals during gait (Cuesta-Vargas et al. 2010; Fusca et al. 2018). Mundt et al extended their training dataset to include both simulated and measured IMU data to estimate joint kinematic profiles using artificial NNs (Mundt et al. 2021). Dorschky et al. also found that the addition of synthetic IMU data improved their model predictions (Mundt et al. 2020b). However, these studies were restricted to walking and treadmill activities. Recently , Tan et al. implemented a Bidirectional LSTM (BiLSTM) model to estimate joint kinematics in the sagittal plane using IMUs for osteoarthritis (OA) patients performing activities of daily living: gait, sit-to-stand, and negotiating stairs. Hossain et al. achieved a low error rate in lower extremity joint kinematic predictions using feet IMUs across level walking, treadmill, ramp, stair ascent, and stair descent activities with DeepBBWAE-Net, an ensemble CNN-RNN based deep learning model (Dorschky et al. 2020).

A common preprocessing step for deep learning model development is segmentation of kinematics data and the corresponding IMUs signals into individual gait cycles with a consistent length achieved by normalizing the data with respect to time (Hossain et al.

113

2022). Most previous studies used MOCAP data for segmentation. In a novel approach, Mundt et al. (2020) predicted joint kinematics based on a continuous stream of IMU data without prior segmentation. They then compared the performance of the LSTM model trained on a longer motion sequence against time-normalized gait cycles and found that a longer motion sequence resulted in superior performance (Dorschky et al. 2020; Hernandez et al. 2021; Hossain et al. 2022; Mundt et al. 2020b, 2020a; Renani et al. 2021). Hernandez et al. (2021) also utilized a continuous time series for training their convolutional neural network long short term memory model (CNNLSTM) but studied its performance in other activities such as walking, running, and gait transition (Mundt et al. 2020a).

In practical applications where subjects wear only IMUs, segmentation of cycles with kinematic data is not feasible. One possible solution is segmentation based on the characteristics of the IMU data. Proposed methods to segment IMU data are currently limited to gait activities and healthy populations (Hernandez et al. 2021). The feasibility of these methods for applications in complex activities of daily living, such as transitioning from gait to stair ascent, turning, sitting to walking, etc., has not been fully investigated. These methods may not apply to individuals with musculoskeletal pathology as they produce abnormal movement patterns (Celik et al. 2021; Rampp et al. 2014; Romijnders et al. 2022). The time and computational cost of the additional preprocessing steps required for continuous real-time joint kinematic estimation reduce the desirability of this approach.

While progress has been made, advancements in machine learning methods in biomechanics remain comparably slow to similar applications in language processing and image recognition. One impeding factor is the lack of publicly available datasets, source

codes, and models. This limits the development and evaluation of models to only a small group of researchers and delays progress. Publicly available models would allow for additional opportunities to implement state-of-the-art machine learning techniques such as transfer learning (Al-Zahrani and Bakheit 2002; Henriksen et al. 2010; Szopa et al. 2020), fine-tuning, or one-shot and zero-shot learning. Thanks to Camargo et al. (2021), a publicly available dataset has been introduced containing 3-dimensional kinematics and wearable sensor data from 22 adults for multiple locomotion tasks including level walking (LW), ramp ascent (RA), ramp descent (RD), stair ascent (SA), and ramp descent (SD). Using such datasets, machine learning models for various applications can be developed and there is a greater opportunity for researchers to advance the field (Goodfellow et al. 2016). As of this time, there are no publicly available kinematic-prediction models.

Machine learning models that have been used in previous studies were limited to NN models, including RNN, CNN, LSTM, and fully connected NNs in various combinations (Camargo et al. 2021). These models provided reliable performance in mapping IMU signals to joint kinematics. However, recent research in the field of deep learning has shown that a relatively new model, the transformer, outperformed previous models in many tasks and is increasingly the model of choice for solving deep learning problems. The transformer was introduced in 2017 by a team at Google Brain for natural language processing tasks to overcome the limitations of RNNs for sequence data. RNNs have difficulty capturing long-term dependencies and processing sequential data in parallel. A transformer, on the other hand, uses self-attention to capture global dependencies while processing sequences in parallel (Dorschky et al. 2020; Hernandez et al. 2021; Hossain et

al. 2022; Mundt et al. 2019, 2020a, 2021; Renani et al. 2021; Tan et al. 2022). Transformers have evolved beyond language tasks into other areas such as time series analysis (Vaswani et al. 2017) and computer vision (Sun et al. 2021; Wu et al. 2020; Zerveas et al. 2020; Zhou et al. 2020a). The potential of this model in biomechanics tasks has not yet been investigated.

To explore the use of transformer models in biomechanics applications, the current study has three aims. The first aim is to implement transformer-based models for predicting joint kinematics from continuous streams of unsegmented IMU signals across gait, ramp, and stair activities. The second aim is to compare the performance of transformer-based models against previous models such as BiLSTM and CNNLSTM. The final aim is to investigate whether a single universal model for all activities has superior performance compared to activity-specific models. The resulting trained models will be open source, enabling studies of reproducibility and the advancement of the field. We hypothesize that (1) the transformer-based model will outperform other models in predicting joint kinematics and (2) activity-specific models will perform equivalent to models trained across all activities.

6.3 Methods

6.3.1 Dataset

A publicly available lower limb biomechanics dataset has been used in this study (Han et al. 2022; Ruan et al. 2022; Zerveas et al. 2020). This comprehensive dataset includes IMU data along with the kinematic and kinetic profiles of joint biomechanics from 19 healthy subjects performing LW, RA, RD, SA, and SD. Each subject was outfitted

unilaterally on the right side with 4 six-axis IMUs (Yost, Ohio, USA), and bilaterally with 32 motion capture markers (Vicon. Ltd., Oxford, UK). IMUs were attached to the anterior surface of the foot, shank, and thigh at ¾ of the length of each segment and the anterior surface of the torso between the sternum and navel. Ground reaction forces were also recorded using force plates (Bertec, Ohio, USA) located in the instrumented treadmill and level with the floor, ramp, and stairs. Joint kinematics and kinetics were calculated by analyzing the MOCAP data along with ground reaction forces using inverse kinematics and inverse dynamics in OpenSim (Camargo et al. 2021). The current study utilized data from the IMUs on the lower limb and sagittal plane joint kinematics at the hip, knee, and ankle from 19 subjects across five activities, including LW at three self-selected speeds, RA, RD, SA, and SD. The number of samples for each group was 1170, 1204, 1204, 789, and 789, respectively.

6.3.2 Preprocessing

IMU and kinematic data were down-sampled from 200 Hz to 100 Hz. The data were arbitrarily segmented into samples of 256 points using a sliding window with 50% overlap per trial. A zero-padding technique was used to ensure the data was a consistent length prior to use in the deep-learning models. The length of 256 was selected to ensure that each sample consisted of at least two successive gait cycles and limitations of the graphic processing units (GPUs) used during training and evaluation. The IMU data were scaled using the standardization method to facilitate gradient descent convergence during training (Goodfellow et al. 2016). A total of 2523, 3369, 3491, 1451, and 1258 samples were generated for LW, RA, RD, SA, and SD, respectively.

6.3.3 Neural Network Models

Three conventional deep NN models (BiLSTM, CNNLSTM, and a new biomechanical multi-activity transformer-based model called BioMAT) and a multi-linear regression (MLR) model were used for mapping IMU data to hip, knee, and ankle flexion kinematics.

6.3.3.1 Multi-Linear Regression Model

A MLR model was chosen as the baseline for this study. Input data to the MLR was reshaped from 3D [number of cycles, length of time series, IMU channels] to 2D [number of cycles, length of time series x IMU channels] and then reshaped from 2D to 3D after kinematic predictions. The MLR model included coefficients that were optimized by minimizing the residual sum of squares between the measured and predicted joint kinematics.

6.3.3.2 CNNLSTM Architecture

CNNLSTM is an architecture specifically designed for sequence prediction with spatial inputs like images or videos. CNNLSTMs consist of multiple convolutional layers, followed by multiple LSTM layers and a final dense or fully connected layer (Delp et al. 2007). Feature extraction occurs with convolutional layers (spatial domain) while time-series prediction is accomplished with recurrent layers (time domain). This model has been used for activity recognition and joint kinematic predictions in previous studies (Bao et al. 2019; Hernandez et al. 2020, 2021; Ordóñez and Roggen 2016). The current study implemented a Deep CNNLSTM based on Hernandez et al. (Bao et al. 2019; Hernandez et al. 2020, 2021; Ordóñez and Roggen 2016) with two 2D CNN layers followed by two LSTM layers BiLSTM is a type of recurrent neural network, which is a class of neural

network effective in time series regression tasks that temporally propagates information with each new estimate. As opposed to unidirectional LSTM models which only consider information from the past, BiLSTM models also consider information from future inputs to improve accuracy. The performance of BiLSTM was demonstrated in a similar studies (Hernandez et al. 2021). LSTMs mitigate the vanishing gradient problem prevalent in RNNs with a gated structure and cell state within each node. The BiLSTM used in this study was composed of two LSTM layers of size 50 and a fully connected layer that reshaped the network output to one size (Fig. 6.1b) (Renani et al. 2021; Tan et al. 2022).

6.3.3.3 BioMAT Architecture

Transformer models operate based on an attention mechanism. The original motivation behind developing transformer models was to solve natural transduction or language translation problems (Renani et al. 2021; Tan et al. 2022). This model is ideal for sequence-to-sequence mapping (Zerveas et al. 2020). Given the current study is related to mapping a sequence of IMU data to a sequence of joint kinematics, as well as the reliable performance of transformers in applications such as forecasting, object detection, and computer vision tasks, transformer models are an ideal candidate.

Transformer models can consist of an encoder and a decoder, which are connected by an attention layer. The encoder maps the input sequence to a vector representation. The decoder generates the output sequence from that vector representation. Bidirectional encoder representations from transformers (BERT) and generative pre-trained transformers (GPT) are two well-known systems that have been trained on large databases. BERT only includes an encoder and is typically trained using supervised learning for tasks such as text

119

classification or named entity recognition (Devlin et al. 2018). GPT includes both an encoder and decoder and is trained using unsupervised learning. During training, the model learns to predict the next word or number in a sequence based on the previous context (Vaswani et al. 2017). BioMAT was based on the BERT architecture with an encoder consisting of an embedding layer, a positional layer, and a stack of encoder layers each with multi-head attention layers followed by a fully connected layer (Fig. 6.1c). Three additional fully connected layers were added to map the resultant vector from the encoder to three kinematics times series (hip, knee, and ankle). The transformer model utilized in this study was adopted based on previously published work (Oguiza 2022).
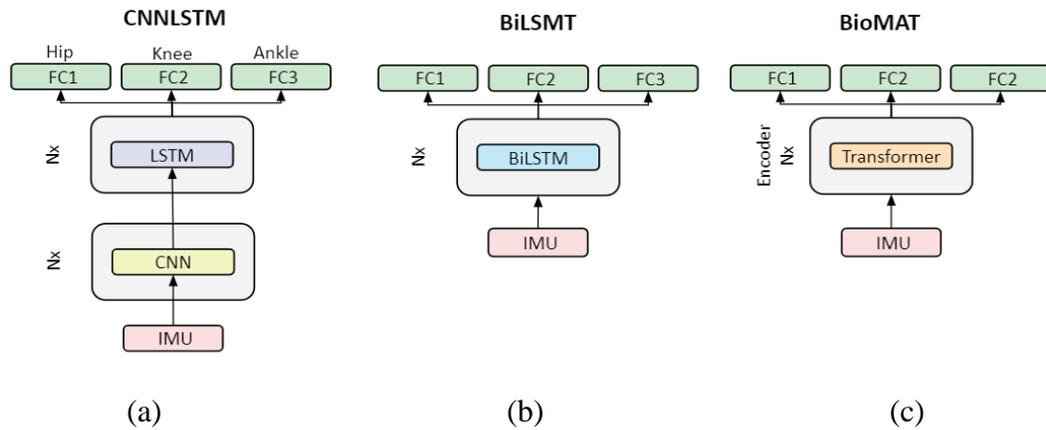
Figure 6.1: Machine learning model architecture: a) CNNLSTM. B) BiLSTM. C) BioMAT.

6.3.4 Training and Parameter Tuning

Data from the 19 subjects were randomly divided into training (16 subjects) and testing (3 subjects) sets. Model training was conducted using adaptive learning rate optimization with a learning rate of 0.001, batch size of 50, and 50 epochs. The cost function used for training was the mean square error between predicted and measured kinematics. An L2

120

regularization coefficient of lambda = 0.001 was used to prevent overfitting. The models were created using PyTorch 1.8.1 in Python 3.7 and trained and evaluated on NVIDIA TiTAN XP GPUs with 12 GB of memory. Hyperparameters for the BioMAT model were selected after tuning with a 5-fold cross-validation by subjects on training data across all five activities or selected based on previously published studies (Mundt et al. 2020a; Renani et al. 2021). Table 6.1 includes the list of hyperparameters for each model.

Table 6.1: Selected hyperparameters for each model

CNNLSTM (Tan et al. 2022)	BiLSTM (Hernandez et al. 2021)	BioMAT (Ours)
CNN2D-1 kernel size: 10,3 CNN2D-1 n output: 16 CNN2D-2 kernel size: 10,3 CNN2D-2 n output: 32 LSTM hidden size: 128 LSTM n layers: 2 dropout: 0.2	BiLSTM hidden size: 128 BiLSTM n layers: 2 dropout: 0.2	BioMAT d model: 256 BioMAT n heads: 16 BioMAT d ff: 128 BioMAT n layers: 4 res dropout: 0.5 fc dropout: 0.5

BioMAT d model: Total dimension of the model (number of features created by the model)
BioMAT n heads: Parallel attention heads
BioMAT d ff: The dimension of the feedforward network model
res dropout: Amount of residual dropout applied in the encoder
fc dropout: Dropout applied to the final fully connected layer

6.3.5 Neural Network Evaluation and Statistical Test

To investigate the generalizability of the models in predicting joint kinematics across various activities, the performance metrics of each model, after training on the combined dataset of all activities, were reported for predictions of each individual activity and predictions across all activities combined. The performance metrics included root mean square error (RMSE), normalized root mean square error (nRMSE), and Pearson correlation coefficient (r) between measured and predicted kinematics.

In a subsequent analysis, the models were re-trained for each activity separately (e.g. trained only on gait) and the predictions tested on that same activity as well as for activities not included in the training set (e.g. model trained on gait predicting stair ascent). The same evaluation metrics were used to assess the impact of activity diversity in the training datasets. Errors for tasks were aggregated by taking the mean across all joints of a specific activity and test subjects. A two-way multivariate analysis (MANOVA) was conducted to detect interactions between the two independent variables, training activities and test activities, and RMSE and r for BioMAT. A multiple comparison test was also conducted to compute pairwise differences between models trained on different training activities for each test activity.

6.4 Results

All machine learning models evaluated in the study produced reasonable joint kinematic predictions for each activity after training simultaneously on all activities (Fig. 6.2). BioMAT achieved lower RMSE and nRMSE across all three joints compared to BiLSTM, CNNLSTM, and MLR models (Table 6.2). BioMAT, BiLSTM, CNNLSTM, and MLR models achieved RMSE of $5.5 \pm 0.5°$, $7.0 \pm 1.0°$, $8.8 \pm 2.3°$, and $14.1 \pm 7.3°$ for joint kinematics predictions, respectively. A similar trend was observed for nRMSE with the smallest nRMSE of 5.4 ± 1.2 and the largest nRMSE of 24.2 ± 12.7 for BioMAT and MLR models, respectively. The mean correlation coefficients between model predictions and measured kinematics ranged from 0.91 ± 0.04 to 0.98 ± 0.01 for the MLR model at ankle joint and BioMAT at the knee, respectively.

When trained on specific activities, BioMAT likewise demonstrated the lowest RMSE and nRMSE among model architectures for all five tasks with average RMSE and nRMSE of 5.5 ± 1.1° and 6.8 ± 1.6°, respectively. BioMAT yielded similar correlation coefficients to BiLSTM for LW (0.97 ± 0.03), RA (0.97 ± 0.02), and SD (0.98 ± 0.02), and was slightly higher for RD (0.94 ± 0.02) and lower for SA (0.97 ± 0.04) (Table 6.3).

Increased activity diversity in the training set improved prediction accuracy for certain model architectures. For example, training the BioMAT and CNNLSTM architectures simultaneously on all activities improved prediction accuracy for four out of five activities compared to training on a specific activity (Tables 6.3-4). However, the post hoc multiple comparison tests for the BioMAT models indicated the accuracy differences were not statistically significant. Conversely, training the MLR and BiLSTM architectures simultaneously on all activities reduced the prediction accuracy for three out of five activities. The two-way MANOVA identified significant main and interaction effects among the type of training data (all activities versus activity-specific) and test activity for both RMSE and r in the BioMAT model (F(50,142) = 2.674, p ≤.001, Wilks' Λ = 0.265). As expected, statistically significant reductions in accuracy (RMSE and r) were observed when the activity-specific models were used to predict kinematics from other activities (Figure 6.3).

Table 6.2: RMSE, nRMSE, and r (mean ± standard deviation) between model predictions and ground truth kinematics for models trained on all activities simultaneously across all subjects in the test set. Bold indicates most accurate model architecture for that joint metric.

Metrics	Joint	Hip	Knee	Ankle	Mean
RMSE (°)	MLR	20.3 ± 11.8	10.1 ± 1.9	11.9 ± 8.3	14.1 ± 7.3
	CNNLSTM	10.9 ± 2.2	10.5 ± 3.9	5.09 ± 0.8	8.8 ± 2.3
	BiLSTM	9.2 ± 1.4	6.9 ± 1.1	4.8 ± 0.8	7.0 ± 1.0
	BioMAT	**6.4 ± 1.0**	**5.5 ± 1.1**	**4.6 ± 0.7**	**5.5 ± 0.5**
nRMSE	MLR	24.2 ± 12.7	10.0 ± 2.3	17.3 ± 10.0	17.2 ± 7.8
	CNNLSTM	13.5 ± 3.5	10.6 ± 4.7	8.1 ± 2.7	10.7 ± 3.2
	BiLSTM	11.6 ± 3.4	6.8 ± 1.0	7.5 ± 1.8	8.6 ± 1.0
	BioMAT	**7.9 ± 1.6**	**5.4 ± 1.2**	**7.1 ± 0.9**	**6.8 ± 0.3**
r	MLR	0.92 ± 0.06	0.95 ± 0.04	0.91 ± 0.04	0.92 ± 0.04
	CNNLSTM	0.92 ± 0.04	0.93 ± 0.06	0.91 ± 0.07	0.92 ± 0.05
	BiLSTM	**0.97 ± 0.04**	0.98 ± 0.02	**0.96 ± 0.02**	**0.97 ± 0.02**
	BioMAT	**0.97 ± 0.03**	**0.98 ± 0.01**	0.95 ± 0.02	0.96 ± 0.01
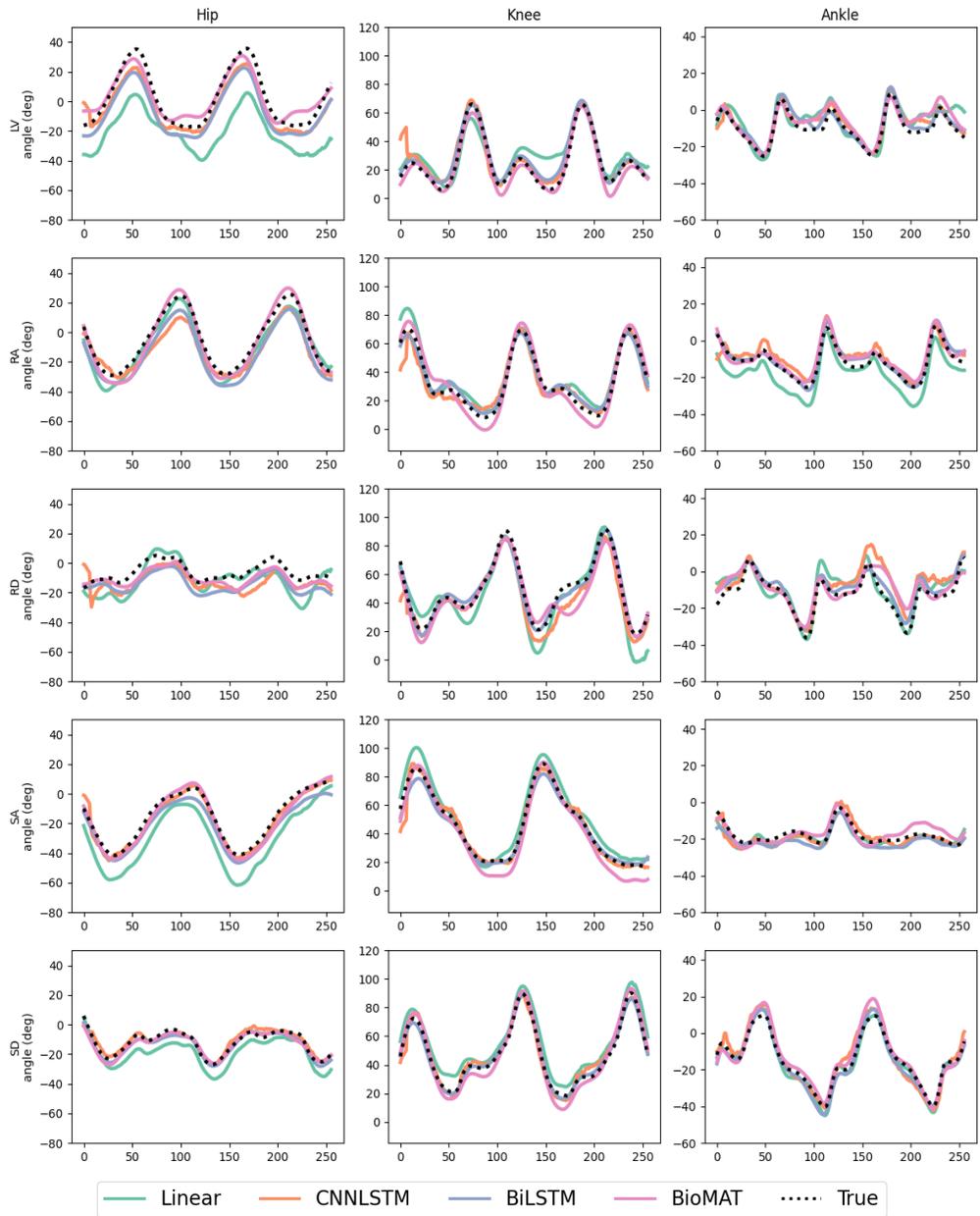
Figure 6.2: Representative ground truth and predicted joint kinematics across different activities and for a test subject from models trained simultaneously on all activities. Ground truth (dash line) and prediction (solid) for different predictive models (LW: Level Walking, RA: Ramp Ascent, RD: Ramp Descent, SA: Stair Ascent, SD: Stair Descent).

Table 6.3: RMSE, nRMSE, and r (mean ± standard deviation) between model predictions and ground truth kinematics for models trained on all activities and tested on individual activities. Bold indicates most accurate model architecture for that activity. (LW: Level Walking, RA: Ramp Ascent, RD: Ramp Descent, SA: Stair Ascent, SD: Stair Descent)

Metric	Model	Train: All Test: LW	Train: All Test: RA	Train: All Test: RD	Train: All Test: SA	Train: All Test: SD
RMSE°	MLR	8.5 ± 2.1	21.7 ± 10.3	22.5 ± 10.8	8.9 ± 3.4	9.0 ± 3.5
	CNNLSTM	12.3 ± 5.6	9.7 ± 3.8	7.8 ± 2.4	6.8 ± 2.3	7.5 ± 2.7
	BiLSTM	5.3 ± 1.6	7.5 ± 2.1	7.4 ± 2.1	7.5 ± 2.6	7.3 ± 2.8
	BioMAT	**5.0 ± 1.5**	**6.2 ± 1.1**	**5.8 ± 1.1**	**5.3 ± 1.6**	**5.2 ± 0.7**
nRMSE	MLR	11.8 ± 1.7	23.3 ± 9.4	27.7 ± 15.9	10.9 ± 3.0	12.3 ± 8.2
	CNNLSTM	16.3 ± 3.3	10.1 ± 2.1	9.2 ± 2.9	8.3 ± 1.6	9.8 ± 5.7
	BiLSTM	7.3 ± 1.7	8.0 ± 0.4	8.7 ± 2.8	9.4 ± 2.8	9.7 ± 6.3
	BioMAT	**7.2 ± 2.4**	**6.7 ± 0.4**	**6.7 ± 0.3**	**6.6 ± 1.4**	**6.8 ± 3.0**
r	MLR	0.92 ± 0.03	0.92 ± 0.04	0.87 ± 0.05	0.96 ± 0.04	0.95 ± 0.05
	CNNLSTM	0.85 ± 0.04	0.9 ± 0.04	0.92 ± 0.02	0.97 ± 0.03	0.95 ± 0.03
	BiLSTM	**0.97 ± 0.02**	**0.97 ± 0.01**	0.93 ± 0.02	**0.98 ± 0.02**	**0.98 ± 0.01**
	BioMAT	0.97 ± 0.03	0.97 ± 0.02	**0.94 ± 0.02**	0.97 ± 0.04	0.98 ± 0.02

Table 6.4: RMSE, nRMSE, and r (mean ± standard deviation) between model predictions and ground truth kinematics for models trained on a single activity and tested on that same activity. Bold indicates most accurate model architecture for that activity.

(LW: Level Walking, RA: Ramp Ascent, RD: Ramp Descent, SA: Stair Ascent, SD: Stair Descent).

Metric	Model	Train: LW Test: LW	Train: RA Test: RA	Train: RD Test: RD	Train: SA Test: SA	Train: SD Test: SD
RMSE°	MLR	9.6 ± 3.5°	31.2 ± 10.6°	13.8 ± 2.4°	7.9 ± 3.5°	7.9 ± 1.3°
	CNNLSTM	6.2 ± 2.2°	10.3 ± 4.5°	8.3 ± 1.4°	13.4 ± 5.2°	18.8 ± 6.8°
	BiLSTM	5.5 ± 1.6°	8.2 ± 2.9°	7.0 ± 2.0°	5.3 ± 1.7°	7.2 ± 2.1°
	BioMAT	**5.3 ± 2.1°**	**6.7 ± 2.0°**	**6.9 ± 2.2°**	**4.9 ± 1.4°**	**5.6 ± 1.3°**
nRMSE	MLR	13.1 ± 3.2	33.2 ± 6.7	16.2 ± 2.2	9.5 ± 2.7	10.1 ± 3.6
	CNNLSTM	8.4 ± 1.6	10.6 ± 2.3	9.7 ± 1.2	16.0 ± 1.5	22.4 ± 0.9

------------------	BiLSTM	7.5 ± 0.3	8.6 ± 1.2	**8.2 ± 2.4**	6.5 ± 1.2	9.5 ± 5.1
	BioMAT	**7.3 ± 2.3**	**7.1 ± 0.6**	8.2 ± 3.1	**5.9 ± 0.7**	**7.5 ± 3.9**
	MLR	0.91 ± 0.04	0.90 ± 0.03	0.83 ± 0.06	0.96 ± 0.04	0.95 ± 0.02
	CNNLSTM	0.94 ± 0.03	0.91 ± 0.04	0.88 ± 0.03	0.70 ± 0.30	$-0.0\,2 \pm 0.04$
r	BiLSTM	0.97 ± 0.03	0.96 ± 0.01	0.93 ± 0.03	0.98 ± 0.03	0.98 ± 0.01
	BioMAT	**0.97 ± 0.02**	**0.97 ± 0.01**	**0.95 ± 0.03**	**0.97 ± 0.03**	**0.98 ± 0.02**

6.5 Discussion

This study introduced an adaptation of a state-of-the-art transformer-based model (BioMAT) for predicting joint kinematics of lower extremities based on streams of IMU data including acceleration and angular velocity. BioMAT consistently yielded the highest performance with the lowest RMSE, nRMSE, and highest correlation coefficients compared to other published models at all three joints and across all five activities of daily living. When trained with data from all activities of daily living, BioMAT's prediction accuracy was improved compared to training purely on activity-specific data for four out of five tasks. Further, this performance was achieved without the need to segment the input IMU data into discrete gait cycles.

Zerveas et al. introduced a transformer-based model for multivariate time series representation learning in 2020 (Renani et al. 2020). Their modeling approach generated the most accurate method for multivariate time series classification and regression tasks on several benchmark datasets when compared to contemporary models such as XGboost (Krishnapuram et al. 2016) and ResNet (He et al. 2015). Siddhad et al. demonstrated that the transformer model outperformed BiLSTM and CNN models in a study to classify electroencephalograms (Zerveas et al. 2020). These studies built the foundation for the

current transformer-based model for multivariate time series in joint kinematic predictions from IMUs. BioMAT was likewise compared against CNNLSTM (Siddhad et al. 2022) and BiLSTM (Hernandez et al. 2021) architectures, the predominant models in recent literature for joint kinematic predictions, showing compelling results. The proposed BioMAT model demonstrated superior prediction accuracy with an average RMSE of 5.5° across all three joints and activities, compared to BiLSTM and CNNLSTM with average RMSEs of 7.0° and 8.8° respectively. BioMAT also achieved smaller standard deviations in RMSE across all joints and tasks (BioMAT standard deviations: joint level = 0.5° and task level = 1.1°), compared to BiLSTM and CNNLSTM (BiLSTM standard deviations: joint level = 1.0° and task level = 2.1°, and CNNLSTM standard deviations: joint level = 2.3° and task level = 3.7°). The smaller standard deviations demonstrate the increased reliability of the transformer compared to other models.

The machine learning models used in this study improved the prediction accuracy relative to the benchmark MLR by 37% to 61% for RMSE, 69% to 93% for nRMSE, and up to 5% for the correlation coefficient. The hip and ankle joints had the largest and smallest RMSE across all models, respectively. When normalized over the range of the kinematics data, the knee joint achieved the highest nRMSE. This indicated that the deep learning models were most robust for the knee joint, then the ankle, and lastly the hip. The correlation coefficients were consistent across the joints for all models. The highest correlations were observed for the knee and the lowest for the ankle. Earlier studies have also observed this trend (Tan et al. 2022). A plausible explanation for decreased predictive ability at the ankle joint is that the smaller range of motion generates a reduced signal-to-

noise ratio in the IMU measurements (Dorschky et al. 2020; Gholami et al. 2020; Hossain et al. 2022; Renani et al. 2021). When comparing model performance across different tasks, BiLSTM and BioMAT had similar accuracy with the lowest average RMSE of $5.3 \pm 1.6°$ and $5.0 \pm 1.5°$ for LW and the largest RMSE with values of $7.5 \pm 2.1°$ and $6.2 \pm 1.1°$ for RA, respectively. In contrast, the CNN model had its lowest accuracy for LW with an average RMSE of $12.3 \pm 5.6°$, and its best performance in SA with an RMSE of $6.8 \pm 2.3°$.

Mundt et al. evaluated an LSTM model's kinematics predictions during level walking at different speeds using approximately 88,000 simulated IMU samples from 150 subjects and achieved an RMSE of $1.6°$, $1.7°$, and $1.4°$ with r of 0.98, 0.99, and 0.94 across the hip, knee, and ankle joints respectively (Tan et al. 2022). Hernandez et al. employed a CNNLSTM model with 27 subjects performing treadmill walking and running tasks with reported mean absolute errors of $3.8°$, $3.0°$, $4.9°$, and r of 0.99, 0.99, and 0.97 across three joints (Hernandez et al. 2021). Compared to these studies, BioMAT had higher RMSEs (hip = $6.8°$, knee = $4.2°$, ankle = $4.2°$), mean absolute error (hip = $5.5°$, knee = $3.3°$, ankle = $3.3°$), and a lower correlation coefficient for the ankle joint (0.93). This was likely due to a combination of factors, including the larger training sets used in the previous studies, the use of simulated IMU data instead of measured IMU data, and performing walking on a treadmill instead of level ground. Simulated IMU data calculated from the kinematics of a musculoskeletal model does not include noise or skin artifacts inherent in measured IMU data. We have demonstrated in previous work that including synthetic IMU data improves prediction accuracy (Renani et al. 2021). Treadmill walking also provides a more controlled environment compared to walking on level ground resulting in more repeatable

gait patterns (Hossain et al. 2022; Mundt et al. 2020a). Table 6.5 compares the current

results to other similar studies that used measured IMU signals for model training. While

BioMAT demonstrated comparable results, it should be noted that data sets, sensor

positions, numbers of sensors, and environmental conditions varied between studies.

Training models using a public dataset, like the one used in this study, helps to standardize

studies and can facilitate benchmarking various models and methodologies (Hossain et al.

2022).

Table 6.5: Prediction accuracies from previous studies for sagittal lower limb
kinematics. Sensor locations included the pelvis (P), thigh (T), shank (S), and foot (F).
Activities include level walking (LW), level running (LR), treadmill running (TR), ramp
ascent (RA), ramp descent (RD), stair ascent (SA), and stair descent (SD).

				RMSE°			r		
Study	Activity	Model	Sensors	Hip	Knee	Ankle	Hip	Knee	Ankle
Dorschkey et al.2020	LW+LR	2DCNN	PTSF	5.4	5.2	5.5	0.97	0.99	0.96
Gholami et al.2020	TR	1DCNN	F	5.6	6.5	4.7	0.84	0.93	0.78
Tan et al.2022	LW	BiLSTM	TS	NA	8.4	NA	NA	0.85	NA
Tan et al.2022	SA	BiLSTM	TS	NA	9.7	NA	NA	0.95	NA
Tan et al.2022	SD	BiLSTM	TS	NA	10.0	NA	NA	0.86	NA
Sharifi et al.2021	LW	BiLSTM	PTSF	7.2	2.9	NA	0.88	0.99	NA
Hossain et al.2022	LW	DeepBBWAVE-Net	FF	4.3	4.3	3.1	0.97	0.99	0.95
Hossain et al.2022	RA	DeepBBWAVE-Net	FF	5.7	5.0	3.5	0.98	0.98	0.96
Hossain et al.2022	RD	DeepBBWAVE-Net	FF	4.3	6.1	3.7	0.93	0.97	0.94
Hossain et al.2022	SA	DeepBBWAVE-Net	FF	6.0	5.9	4.0	0.98	0.99	0.96
Hossain et al.2022	SD	DeepBBWAVE-Net	FF	5.3	6.8	5.0	0.93	0.97	0.98
Current	LW	BioMAT	TSF	6.8	4.2	4.2	0.99	0.99	0.93
Current	RA	BioMAT	TSF	7.3	6.2	5.1	0.98	0.97	0.95

Current	RD	BioMAT	TSF	4.9	7.0	5.5	0.92	0.97	0.94
Current	SA	BioMAT	TSF	6.9	5.3	3.7	0.99	0.99	0.93
Current	SD	BioMAT	TSF	56.0	4.8	4.7	0.96	0.99	0.98

Deploying machine learning models for kinematic predictions during real-world applications requires several practical steps that are enabled using transformer models. Both BioMAT and CNNLSTM showed similar RMSE values between models trained on all activities versus specific activities. This improves the generalizability of the tool and removes the need for activity classification and gait cycle segmentation prior to kinematic predictions. In addition to the gains in prediction accuracy, BioMAT required less training and inference time for kinematic predictions (Table 6.6). Specifically, the inference time from BioMAT was 0.003 seconds / batch, 79% faster than predictions from the BiLSTM model (Vaswani et al. 2017; Zerveas et al. 2020).

Table 6.6: Number of parameters, training time, and inference time for each model.

Model	# Parameters	Training time (sec / epoch)	Inference time (sec / batch)
BiLSTM	106,635,584	14.2	0.014
CNNLSTM	1,201,046	15.9	0.006
BioMAT	51,257,603	12.9	0.003

There were multiple limitations to this study. First, the dataset used in this study included multiple configurations of stair height (four heights: 102 mm, 127 mm, 152 mm, 178 mm) and ramp inclination angles (6 inclination angles of 5.2°, 7.8°, 9.2°, 11°, 12.4°, and 18°) as well as different speeds for level walking. Including greater variability in the training dataset likely improved the models' generalizability, however, the effect of each configuration on the models' performance was not examined and was outside the scope of the current study. Second, the sensitivity of each model's performance to the number of

data points in the inputs (e.g. 256) has not been investigated. Using a longer sequence length may further improve the models' accuracy (Mundt et al. 2020a; Renani et al. 2021). Third, the hyperparameters associated with the CNNLSTM and BiLSTM were selected based on previously reported studies to allow a direct comparison to the published results. It is plausible that hyperparameter tuning may improve the prediction accuracy of these models. Fourth, the current study focused on evaluating the performance of BioMAT for activity-specific and multi-activity training sets. but the contribution of each training activity to the final model performance remains unclear. Finally, the current study was limited to predicting joint angles in the sagittal plane and the accuracy of the proposed model for predicting joint angles in the coronal and axial planes has not been evaluated. Since the current model and dataset are both open-sourced, future researchers could leverage the current method and model to address some of these limitations.

In future work, the current model could be used to investigate machine learning techniques such as transfer learning (Zhou et al. 2022; Zhuang et al. 2019), fine-tuning, and one- or zero-shot learning methods (Rezaei and Shahidi 2020; Xian et al. 2017) for relevant biomechanical tasks or datasets. Transfer learning is a powerful technique to achieve highly accurate results on a wide range of tasks (Tan et al. 2018; Zoph et al. 2016). In biomechanics, kinematic prediction models trained on one dataset (e.g. gait activities in a healthy population) could be used with transfer learning to evaluate a new task or patient population (e.g. stair ascent in the OA population). Although BioMAT was trained on a healthy population, the tool could be fine-tuned to OA and total joint arthroplasty populations by adding a small number of observations from those groups. This would

reduce the need to collect large quantities of data, preprocess samples, and train models from scratch, saving time, money, and computational resources (Tan et al. 2018). Finally, knowledge distillation could potentially be used to compress the current model to a smaller model without significant loss in performance, improving computational efficiency for deployment on edge devices such as smartphones with limited hardware and resources (Hinton et al. 2015).

6.6 Conclusion

In this study, we evaluated a deep learning transformer architecture, BioMAT, to estimate lower extremity kinematics from a continuous stream of IMU data for multiple activities of daily living. This model was trained using a publicly available dataset. BioMAT predicted joint kinematics with equivalent or lower errors than conventional deep NN models without the additional computational steps associated with activity classification and segmentation of gait cycles. This comprehensive analysis revealed that training the model on a diversity of activities outperformed models trained on specific activities in four out of five tasks. A system equipped with a single multifunction model relying on streams of IMU data can bridge the gap to real-time applications of wearable sensors for monitoring movement in clinical and commercial applications. BioMAT has been made open-source and can be found at the link below. Publicly available datasets and model offer valuable resources for other researchers to accelerate advancements in the biomechanics field.

BioMAT is available at: https://digitalcommons.du.edu/biomat/

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusion

Recent advances in wearable sensors and machine learning offer a quantifiable assessment of patient movement that could revolutionize the diagnosis and treatment of movement disorders in the aging population. The purpose of this study was to overcome the technical challenges associated with the use of wearables as a diagnostic tool for osteoarthritic and total joint replacement patients through a detailed biomechanical analysis and the development of machine learning algorithms. To achieve that goal, several sub-goals based on machine learning techniques were defined, implemented, and evaluated. *Spatial-temporal parameter estimation, joint kinematic estimation*, *in-depth analysis of OA and TKA joint kinematics*, and *last but not least develop a single multi-purpose model for joint kinematic estimation.*

Chapter 3 described deep learning-based methods to process a stream of IMUs data and estimate the 12 spatial-temporal gait parameters associated with OA and TAK patients. In this study, the ability of multiple deep neural networks for this task was benchmarked. Additionally, using a comprehensive analysis of various sensor combinations and their sensitivity to STGPs, patient population, and walking pace were investigated. As a result of our research, we showed that deep learning can bypass the dependence on sensor location which makes it difficult to design patient monitoring systems and negatively

impacts patient compliance. More specifically, the 1D convolution neural network (CNN) architecture proposed by Zrenner et indicated the height performance compared to other competitive architectures. This model included two convolutional layers followed by two max-pooling layers, a flattening layer, and two fully-connected layers. In terms of model performance, the percent error ranged from 2.1% (stride time) to 73.7% (toe-out angle) across the 12 STGPs. Overall, however, it was more accurate for temporal parameters than spatial parameters. Regarding sensor combinations, feet-thighs and singular pelvis proved to be the most and least accurate. Additionally, we demonstrated the importance of sufficient variability in training and test data as well as considering the training data set completeness and in-distribution generalization prior to interpreting prediction accuracy and model deployment, especially for clinically relevant data with small sample sizes.

Chapter 4 focused on the development of a deep-learning framework to translate the stream of IMU data into joint kinematics of patients with OA and TKA. Additionally, this study proposed a novel method to overcome the bottlenecks in deep learning model development which happens due to the lack of sufficient training data and the significant time and resources necessary to acquire these datasets. Using this methodology, synthetic kinematics and the associated predicted IMU signals were generated using open-source musculoskeletal modeling software for training joint-specific deep learning models. The validity of this method was assessed through comparisons of generated synthetic IMU data against the measured data. Models trained using the synthetic data outperformed models using only the measured data in five of the six rotational degrees of freedom at the hip and knee. Moreover, when models were trained on both measured and synthetic data, root mean

square errors were reduced by 54% at the hip (RMSE: 1.9°) and 45% at the knee (RMSE: 1.7°), compared to measured data alone. As a result of these findings, future models can be developed for a variety of clinical activities without the burden of using large amounts of gait lab data for model training, streamlining model development, and ultimately improving model performance.

Chapter 5 described biomechanical gait and stair motion characteristics between knee OA patients and TKA patients using principle component analysis PCA, with the notion that stair motion would highlight locomotor strategies not present in walking due to its more effortful nature. Therefore, these motion deviations would make the pathological condition more obvious. Principal component models were created for the lower extremity joint kinematics at each activity about the three axes of the pelvis, three axes of the hip, and axes of the knee and ankle. As a result of these models, PC1 stair descent showed the highest variance between the OA and TKA groups, followed by PC2 and PC3 of gait, and PC4 stair ascent activities. Our finding provides clear evidence that OA and TKA patients can be more distinguishable when they are being analyzed in more high-demand activities, such as stair descent. Our results indicated that considering the whole lower extremity joint or even the whole body in the analysis can produce more reliable outcomes, especially when we are dealing with low sample data.

Additionally, a PCA-based autoencoder was introduced to generate conditional synthetic kinematic data for each patient group and activity. Our results indicate that by increasing the number of synthetic samples, the waveform of synthetic kinematic data will approach the waveform of original kinematics. This method can be leveraged to use in

future studies for generating realistic synthetic kinematic data from existing kinematic profiles for different patient populations. Additionally, through manual tweaking of each PC (e.g. increasing the PC2 of gait which leads to increasing the knee joint angle), we can produce desirable kinematics while preserving the functional relevance of the motion. Other advantages of this method are bypassing the need for large-scale data collection for each population and protecting patient privacy in which the synthetic data can be used for training machine learning-based models instead of actual patient data.

Chapter 6 focused on introducing a single multipurpose transformer-based model, BioMAT, that uses a continuous stream of IMUs data to predict the joint kinematics across various activities of daily living such as level walking, ramp ascent, ramp descent, stair ascent, and stair descent. Unlike previous models, the proposed model is equipped to receive any arbitrary length of IMUs and estimate the joint kinematic accordingly without the need for segmentation of IMUs to a sequence of gait cycles or classifying the activities prior to feeding into the model. In this way, the computational cost and complexity of the model have been reduced, enabling the model to predict joint kinematics in real-time under any field conditions regardless of the activity. The BioMAT indicated the lowest RMSE and nRMSE across three joints and five activities, and when it was compared to other conventional machine learning models, Linear Regression, CNNLSTM, and BiLSTM models. The BiLSTM model ranked 2nd followed by CNNLSTM and Linear Regression model. Finally, training the model on activities verse single activity indicated a mix of results across the models and activities. For BioMAT and CNNLSTM, training on all activities outperformed the model training on a single activity across 4 out of 5 tasks. While

these values were 3 out of 5 and 1 out of five for BiLSTM and Linear Regression models respectively.

7.2 Recommendations

The studies presented in this dissertation were mainly focused on leveraging machine learning models to translate a stream of wearable sensors, IMUs, into clinically relevant metrics. Ultimately, this project aims to provide a simple clinical and consumer-based tool to track the quality of a patient's movement in the clinic or in their own home environment. It also offers a method to translate those movements into diagnostic metrics for use by clinicians to individualize treatment plans and for use by patients to monitor their disease progress or treatments. Although the goal of the dissertation was achieved, there is always room for improvement in the models, dataset, and methods to provide more comprehensive and accurate results. Below are some of those recommendations.

1) The estimation of spatial-temporal gait parameters from IMUs using machine learning models was limited to only gait activity. Therefore, future studies could extend the presented method across other activities of daily living, such as stairs, ramps, sit-to-stand, lunge, etc.

2) The robustness of the model's prediction to the position and orientation of IMUs attached to the limp has not been investigated. Future studies can explore the influence of sensor position and orientation relative to lower extremities segments on model predictions for spatial-temporal parameters and joint kinematics.

3) In the first study, the divergence of subject 21 from the distribution of subjects used to train the CNNs resulted in poor performance of the model, driving up the

reported error for the OA cohort. One of the main challenges in the use of machine-learning models for real-world applications is when the test subjects are outside of the distribution of training data. As a result, methods regarding out-of-distribution detection (Hossain et al. 2022)or domain generalization (DeVries and Taylor 2018; Hsu et al. 2020; Liang et al. 2017) should be explored in the future for improving the robustness of developer machine learning model specifically for clinical applications. Minimal Detectable Change (MDC) and Minimal Clinically Important Difference (MCID) are terms used in clinical research and healthcare to evaluate the magnitude of change in a particular outcome measure, such as joint angle, that is necessary to indicate a real change in the patient's condition. MDC represents the minimum amount of change in a measurement that can be reliably detected above measurement error. MCDI, on the other hand, refers to the smallest change in the patient outcome or physiological variable that is considered clinically meaningful or important for the patient. This study did not evaluate the MCID value as it is mainly dependent on patient perception and satisfaction. Future research should assess the effectiveness of the proposed methods for determining MCID at different time points post-surgery.

4) In the second study, despite a significant improvement achieved in the prediction accuracy of the model by the inclusion of synthetic data in the training dataset, the generated synthetic kinematic data were not physiologically realistic. Additionally, since we randomly augmented the joint kinematics by inducing variations in both time and magnitude, we were not able to control the profile of generated synthetic

data. Future work should consider controllable generative methods to develop targeted and realistic synthetic data that span the variability in the subject population of interest such as generative adversarial network, variation autoencoder, diffusion model, etc (Wang et al. 2021; Zhou et al. 2022). Synthetic data generation not only allows for improving the models' accuracy and robustness, but also it bypasses the restriction around patient and health data and privacy.

5) As a result of this work, a comprehensive dataset from OA and TKA patients performing various activities of daily living was generated. This dataset includes IMUs, Motion Capture, Force Plate, Patient Anthropometric data, and short surveys regarding knee health and activity level. Although the generated dataset was significantly larger and complete compared to other studies, it was still limited to only 30 subjects performing activities in a lab environment. Therefore, we recommend exploring opportunities to create larger datasets, from various patient populations, performing activities of daily living in a non-control environment.

6) Due to the cumulative nature of research, open datasets have an enormous impact, especially in fields the intersection of machine learning and biomechanics exist. The availability of these resources facilitates the development of new analyses, new models, data practices, and reproducible results. Therefore, we highly encourage researchers to investigate avenues that allow open-sourcing large-scale datasets. Similar claims applied to codes and models. Unfortunately, majorities of previous research failed to publish the trained models or codes which hinders the reproducibility of the work and slows down the advancement of machine learning

in biomechanics. To tackle that, future researchers should consider publishing their codes or model as best practices in their studies.

7) One of the main benefits of open-source machine learning models is to allow investigation of more recent and effective methods in deep learning such as transfer learning, fine-tuning, one-shot or zero-shot learning, and metal learning. The current work has tried to take the initial steps by open-sourcing the models trained on open-source datasets as well as a tool for generating synthetic kinematic data. Future work should consider exploring the aforementioned methods to improve the accuracy and robustness of the models via available open-source models and datasets.

REFERENCES

Abadi, M, Agarwal, A, Barham, P, … Zheng, X (2016) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv. doi:10.48550/arxiv.1603.04467.

Agarwal, A, Miller, S, Hadden, W, … Abboud, R (2019) Comparison of gait kinematics in total and unicondylar knee replacement surgery. The Annals of The Royal College of Surgeons of England, 101(6), 391–398. doi:10.1308/rcsann.2019.0016.

Al-Zahrani, KS, and Bakheit, AMO (2002) A study of the gait characteristics of patients with chronic osteoarthritis of the knee. Disability and Rehabilitation, 24(5), 275–280. doi:10.1080/09638280110087098.

Aminian, K, Robert, Ph, Jequier, E, and Schutz, Y (1994) Estimation of speed and incline of walking using neural network. Conference Proceedings. 10th Anniversary. IMTC/94. Advanced Technologies in I & M. 1994 IEEE Instrumentation and Measurement Technolgy Conference (Cat. No.94CH3424-9), 160–162 vol.1. doi:10.1109/imtc.1994.352073.

Andriacchi, TP, and Alexander, EJ (2000) Studies of human locomotion: past, present and future. Journal of Biomechanics, 33(10), 1217–1224. doi:10.1016/s0021-9290(00)00061-0.

Andriacchi, TP, Andersson, GB, Fermier, RW, Stern, D, and Galante, JO (1980) A study of lower-limb mechanics during stair-climbing. The Journal of Bone and Joint Surgery. American Volume, 62(5), 749–57.

Appelboom, G, Camacho, E, Abraham, ME, … Connolly, ES (2014) Smart wearable body sensors for patient self-assessment and monitoring. Archives of Public Health, 72(1), 28. doi:10.1186/2049-3258-72-28.

Ardestani, MM, Malloy, P, Nam, D, Rosenberg, AG, and Wimmer, MA (2017) TKA patients with unsatisfying knee function show changes in neuromotor synergy pattern but not joint biomechanics. Journal of Electromyography and Kinesiology, 37, 90–100. doi:10.1016/j.jelekin.2017.09.006.

Argent, R, Drummond, S, Remus, A, O'Reilly, M, and Caulfield, B (2019) Evaluating the use of machine learning in the assessment of joint angle using a single inertial sensor. Journal of Rehabilitation and Assistive Technologies Engineering, 6, 2055668319868544. doi:10.1177/2055668319868544.

Asai, Y, Tsutsui, S, Oka, H, … Yoshida, M (2017) Sagittal spino-pelvic alignment in adults: The Wakayama Spine Study. PLoS ONE, 12(6), e0178697. doi:10.1371/journal.pone.0178697.

Atallah, L, Wiik, A, Jones, GG, … Yang, G-Z (2012) Validation of an ear-worn sensor for gait monitoring using a force-plate instrumented treadmill. Gait & Posture, 35(4), 674–676. doi:10.1016/j.gaitpost.2011.11.021.

Bailey, GP, and Harle, R (2014) Assessment of Foot Kinematics During Steady State Running Using a Foot-mounted IMU. Procedia Engineering, 72, 32–37. doi:10.1016/j.proeng.2014.06.009.

Baker, R (2006) Gait analysis methods in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 3(1), 4. doi:10.1186/1743-0003-3-4.

Bao, T, Zaidi, SAR, Xie, S, Yang, P, and Zhang, Z (2019) A CNN-LSTM Hybrid Framework for Wrist Kinematics Estimation Using Surface Electromyography. ArXiv. doi:10.1109/tim.2020.3036654.

Barbour, KE, Helmick, CG, Boring, M, and Brady, TJ (2017) Vital Signs: Prevalence of Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation — United States, 2013–2015. Morbidity and Mortality Weekly Report, 66(9), 246–253. doi:10.15585/mmwr.mm6609e1.

Barth, J, Oberndorfer, C, Pasluosta, C, … Eskofier, BM (2015) Stride Segmentation during Free Walk Movements Using Multi-Dimensional Subsequence Dynamic Time Warping on Inertial Sensor Data. Sensors, 15(3), 6419–6440. doi:10.3390/s150306419.

Begg, RK, Palaniswami, M, and Owen, B (2005) Support vector machines for automated gait classification. IEEE Transactions on Biomedical Engineering, 52(5), 828–838. doi:10.1109/tbme.2005.845241.

Bejek, Z, Paróczai, R, Illyés, Á, and Kiss, RM (2005) The influence of walking speed on gait parameters in healthy people and in patients with osteoarthritis. Knee Surgery, Sports Traumatology, Arthroscopy, 14(7), 612–622. doi:10.1007/s00167-005-0005-6.

Benoit, DL, Ramsey, DK, Lamontagne, M, Xu, L, Wretenberg, P, and Renström, P (2006) Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. Gait & Posture, 24(2), 152–164. doi:10.1016/j.gaitpost.2005.04.012.

Bensalma, F, Mezghani, N, Ouakrim, Y, … Hagemeister, N (2019) A multivariate relationship between the kinematic and clinical parameters of knee osteoarthritis population. BioMedical Engineering OnLine, 18(1), 58. doi:10.1186/s12938-019-0676-8.

Bertoli, M, Cereatti, A, Trojaniello, D, … Croce, UD (2018) Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults. BioMedical Engineering OnLine, 17(1), 58. doi:10.1186/s12938-018-0488-2.

Biggs, PR, Whatling, GM, Wilson, C, and Holt, CA (2019a) Correlations between patient-perceived outcome and objectively-measured biomechanical change following Total Knee Replacement. Gait and Posture, 70(December 2018), 65--70. doi:10.1016/j.gaitpost.2019.02.028.

Biggs, PR, Whatling, GM, Wilson, C, Metcalfe, AJ, and Holt, CA (2019b) Which osteoarthritic gait features recover following total knee replacement surgery? PLoS ONE, 14(1), 1--14. doi:10.1371/journal.pone.0203417.

Biship, and M., C (2017) Pattern recognition and machine learning (information science and statistics).

Bolink, SAAN, Laarhoven, SN van, Lipperts, M, Heyligers, IC, and Grimm, B (2012) Inertial sensor motion analysis of gait, sit–stand transfers and step-up transfers: differentiating knee patients from healthy controls. Physiological Measurement, 33(11), 1947–1958. doi:10.1088/0967-3334/33/11/1947.

Bond-Taylor, S, Leach, A, Long, Y, and Willcocks, CG (2021) Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 1–1. doi:10.1109/tpami.2021.3116668.

Bonnet, V, Mazza, C, Fraisse, P, and Cappozzo, A (2013) Real-time Estimate of Body Kinematics During a Planar Squat Task Using a Single Inertial Measurement Unit. IEEE Transactions on Biomedical Engineering, 60(7), 1920–1926. doi:10.1109/tbme.2013.2245131.

Brandon, SCE, Graham, RB, Almosnino, S, Sadler, EM, Stevenson, JM, and Deluzio, KJ (2013) Interpreting principal components in biomechanics: Representative extremes and single component reconstruction. Journal of Electromyography and Kinesiology, 23(6), 1304–1310. doi:10.1016/j.jelekin.2013.09.010.

Brenneman, EC, and Maly, MR (2018) Identifying changes in gait waveforms following a strengthening intervention for women with knee osteoarthritis using principal components analysis. Gait & Posture, 59, 286–291. doi:10.1016/j.gaitpost.2017.07.006.

Brunner, T, Lauffenburger, J-P, Changey, S, and Basset, M (2015) Magnetometer-Augmented IMU Simulator: In-Depth Elaboration. Sensors, 15(3), 5293–5310. doi:10.3390/s150305293.

Burton, WS, Myers, CA, and Rullkoetter, PJ (2021) Machine learning for rapid estimation of lower extremity muscle and joint loading during activities of daily living. Journal of Biomechanics, 123, 110439. doi:10.1016/j.jbiomech.2021.110439.

Camargo, J, Ramanathan, A, Flanagan, W, and Young, A (2021) A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. Journal of Biomechanics, 119, 110320. doi:10.1016/j.jbiomech.2021.110320.

Cappozzo, A, Croce, UD, Leardini, A, and Chiari, L (2005) Human movement analysis using stereophotogrammetry Part 1: theoretical background. Gait & Posture, 21(2), 186–196. doi:10.1016/j.gaitpost.2004.01.010.

Carcreff, L, Gerber, CN, Paraschiv-Ionescu, A, … Aminian, K (2018) What is the Best Configuration of Wearable Sensors to Measure Spatiotemporal Gait Parameters in Children with Cerebral Palsy? Sensors, 18(2), 394. doi:10.3390/s18020394.

Celik, Y, Stuart, S, Woo, WL, and Godfrey, A (2021) Wearable Inertial Gait Algorithms: Impact of Wear Location and Environment in Healthy and Parkinson's Populations. Sensors (Basel, Switzerland), 21(19), 6476. doi:10.3390/s21196476.

Chau, T (2001) A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods. Gait & Posture, 13(1), 49–66. doi:10.1016/s0966-6362(00)00094-1.

Chehab, EF, Andriacchi, TP, and Favre, J (2017) Speed, age, sex, and body mass index provide a rigorous basis for comparing the kinematic and kinetic profiles of the lower extremity during walking. Journal of Biomechanics, 58, 11–20. doi:10.1016/j.jbiomech.2017.04.014.

Chen, L, Wei, H, and Ferryman, J (2013) A survey of human motion analysis using depth imagery. Pattern Recognition Letters, 34(15), 1995–2006. doi:10.1016/j.patrec.2013.02.006.

Chen, RJ, Lu, MY, Chen, TY, Williamson, DFK, and Mahmood, F (2021) Synthetic data in machine learning for medicine and healthcare. Nature Biomedical Engineering, 5(6), 493–497. doi:10.1038/s41551-021-00751-8.

Choi, A, Jung, H, and Mun, JH (2019) Single Inertial Sensor-Based Neural Networks to Estimate COM-COP Inclination Angle During Walking. Sensors, 19(13), 2974. doi:10.3390/s19132974.

Chollet, F, and others (2015) Keras, Github. Retrieved from https://github.com/keras-team/keras

Clark, RA, Bower, KJ, Mentiplay, BF, Paterson, K, and Pua, Y-H (2013) Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. Journal of Biomechanics, 46(15), 2722–2725. doi:10.1016/j.jbiomech.2013.08.011.

Clary, CW, Fitzpatrick, CK, Maletsky, LP, and Rullkoetter, PJ (2013) The influence of total knee arthroplasty geometry on mid-flexion stability: An experimental and finite element study. Journal of Biomechanics, 46(7), 1351–1357. doi:10.1016/j.jbiomech.2013.01.025.

Cuesta-Vargas, AI, Galán-Mercant, A, and Williams, JM (2010) The use of inertial sensors system for human motion analysis. Physical Therapy Reviews, 15(6), 462–473. doi:10.1179/1743288x11y.0000000006.

Cushion, EJ, Warmenhoven, J, North, JS, and Cleather, DJ (2019) Principal Component Analysis Reveals the Proximal to Distal Pattern in Vertical Jumping Is Governed by Two Functional Degrees of Freedom. Frontiers in Bioengineering and Biotechnology, 7, 193. doi:10.3389/fbioe.2019.00193.

Cutti, AG, Ferrari, A, Garofalo, P, Raggi, M, Cappello, A, and Ferrari, A (2009) 'Outwalk': a protocol for clinical gait analysis based on inertial and magnetic sensors. Medical & Biological Engineering & Computing, 48(1), 17. doi:10.1007/s11517-009-0545-x.

Das, HP, Tran, R, Singh, J, … Spanos, CJ (2021) Conditional Synthetic Data Generation for Robust Machine Learning Applications with Limited Pandemic Data. ArXiv.

Dejnabadi, H, Jolles, BM, and Aminian, K (2005) A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes. IEEE Transactions on Biomedical Engineering, 52(8), 1478–1484. doi:10.1109/tbme.2005.851475.

Delp, SL, Anderson, FC, Arnold, AS, … Thelen, DG (2007) OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement. IEEE Transactions on Biomedical Engineering, 54(11), 1940–1950. doi:10.1109/tbme.2007.901024.

Deluzio, KJ, and Astephen, JL (2007) Biomechanical features of gait waveform data associated with knee osteoarthritis An application of principal component analysis. Gait & Posture, 25(1), 86–93. doi:10.1016/j.gaitpost.2006.01.007.

Deluzio, KJ, Wyss, UP, Costigan, PA, Sorbie, C, and Zee, B (1999) Gait assessment in unicompartmental knee arthroplasty patients: Principal component modelling of gait waveforms and clinical status. Human Movement Science, 18(5), 701–711. doi:10.1016/s0167-9457(99)00030-5.

Deluzio, KJ, Wyss, UP, Zee, B, Costigan, PA, and Serbie, C (1997) Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. Human Movement Science, 16(2–3), 201–217. doi:10.1016/s0167-9457(96)00051-6.

Devlin, J, Chang, M-W, Lee, K, and Toutanova, K (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv.

DeVries, T, and Taylor, GW (2018) Learning Confidence for Out-of-Distribution Detection in Neural Networks. ArXiv. doi:10.48550/arxiv.1802.04865.

DORR, LD, OCHSNER, JL, GRONLEY, J, and PERRY, J (1988) Functional Comparison of Posterior Cruciate Retained versus Cruciate-Sacrificed Total Knee Arthroplasty. Clinical Orthopaedics and Related Research, 236(NA;), 36–43. doi:10.1097/00003086-198811000-00005.

Dorschky, E, Nitschke, M, Martindale, CF, Bogert, AJ van den, Koelewijn, AD, and Eskofier, BM (2020) CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data. Frontiers in Bioengineering and Biotechnology, 8, 604. doi:10.3389/fbioe.2020.00604.

Dorschky, E, Nitschke, M, Seifer, A-K, Bogert, AJ van den, and Eskofier, BM (2019) Estimation of Gait Kinematics and Kinetics from Inertial Sensor Data Using Optimal Control of Musculoskeletal Models. Journal of Biomechanics, 95, 109278. doi:10.1016/j.jbiomech.2019.07.022.

Du, H, Tang, H, Gu, J-M, and Zhou, Y-X (2014) Patient satisfaction after posterior-stabilized total knee arthroplasty: A functional specific analysis. The Knee, 21(4), 866–870. doi:10.1016/j.knee.2014.03.007.

Elkarif, V, Kandel, L, Rand, D, Schwartz, I, Greenberg, A, and Portnoy, S (2020) Kinematics following gait perturbation in adults with knee osteoarthritis: Scheduled versus not scheduled for knee arthroplasty. Gait & Posture, 81, 144–152. doi:10.1016/j.gaitpost.2020.07.021.

Eltoukhy, M, Oh, J, Kuenze, C, and Signorile, J (2016) Improved kinect-based spatiotemporal and kinematic treadmill gait assessment. Gait & Posture, 51, 77–83. doi:10.1016/j.gaitpost.2016.10.001.

Fasel, B, Duc, C, Dadashi, F, … Aminian, K (2017) A wrist sensor and algorithm to determine instantaneous walking cadence and speed in daily life walking. Medical & Biological Engineering & Computing, 55(10), 1773–1785. doi:10.1007/s11517-017-1621-2.

Favre, J, Aissaoui, R, Jolles, BM, Guise, JA de, and Aminian, K (2009) Functional calibration procedure for 3D knee joint angle description using inertial sensors. Journal of Biomechanics, 42(14), 2330–2335. doi:10.1016/j.jbiomech.2009.06.025.

Federolf, PA, Boyer, KA, and Andriacchi, TP (2013) Application of principal component analysis in clinical gait research: Identification of systematic differences between healthy and medial knee-osteoarthritic gait. Journal of Biomechanics, 46(13), 2173–2178. doi:10.1016/j.jbiomech.2013.06.032.

Fellin, RE, Rose, WC, Royer, TD, and Davis, IS (2010) Comparison of methods for kinematic identification of footstrike and toe-off during overground and treadmill running. Journal of Science and Medicine in Sport, 13(6), 646–650. doi:10.1016/j.jsams.2010.03.006.

Ferrari, A, Ginis, P, Hardegger, M, Casamassima, F, Rocchi, L, and Chiari, L (2016) A Mobile Kalman-Filter Based Solution for the Real-Time Estimation of Spatio-Temporal Gait Parameters. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24(7), 764–773. doi:10.1109/tnsre.2015.2457511.

Findlow, A, Goulermas, JY, Nester, C, Howard, D, and Kenney, LPJ (2008) Predicting lower limb joint kinematics using wearable motion sensors. Gait & Posture, 28(1), 120–126. doi:10.1016/j.gaitpost.2007.11.001.

Foxlin, E (2005) Pedestrian Tracking with Shoe-Mounted Inertial Sensors. IEEE Computer Graphics and Applications, 25(6), 38–46. doi:10.1109/mcg.2005.140.

Fransen, M, Nairn, L, Bridgett, L, … Harmer, AR (2017) Post-Acute Rehabilitation After Total Knee Replacement: A Multicenter Randomized Clinical Trial Comparing Long-Term Outcomes. Arthritis Care & Research, 69(2), 192–200. doi:10.1002/acr.23117.

Frick, E, and Rahmatalla, S (2018a) Joint Center Estimation Using Single-Frame Optimization: Part 1: Numerical Simulation. Sensors (Basel, Switzerland), 18(4), 1089. doi:10.3390/s18041089.

Frick, E, and Rahmatalla, S (2018b) Joint Center Estimation Using Single-Frame Optimization: Part 2: Experimentation. Sensors (Basel, Switzerland), 18(8), 2563. doi:10.3390/s18082563.

Fusca, M, Negrini, F, Perego, P, Magoni, L, Molteni, F, and Andreoni, G (2018) Validation of a Wearable IMU System for Gait Analysis: Protocol and Application to a New System. Applied Sciences, 8(7), 1167. doi:10.3390/app8071167.

Gaudreault, N, Mezghani, N, Turcot, K, Hagemeister, N, Boivin, K, and Guise, JA de (2011) Effects of physiotherapy treatment on knee osteoarthritis gait data using principal

component analysis. Clinical Biomechanics, 26(3), 284–291. doi:10.1016/j.clinbiomech.2010.10.004.

Geerse, DJ, Coolen, BH, and Roerdink, M (2015) Kinematic Validation of a Multi-Kinect v2 Instrumented 10-Meter Walkway for Quantitative Gait Assessments. PLoS ONE, 10(10), e0139913. doi:10.1371/journal.pone.0139913.

Ghassemi, NH, Hannink, J, Martindale, CF, … Eskofier, BM (2018) Segmentation of Gait Sequences in Sensor-Based Movement Analysis: A Comparison of Methods in Parkinson's Disease. Sensors, 18(1), 145. doi:10.3390/s18010145.

Gholami, M, Napier, C, and Menon, C (2020) Estimating Lower Extremity Running Gait Kinematics with a Single Accelerometer: A Deep Learning Approach. Sensors, 20(10), 2939. doi:10.3390/s20102939.

Goh, SK, Abbass, HA, Tan, KC, … Li, J (2018) Spatio–Spectral Representation Learning for Electroencephalographic Gait-Pattern Classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(9), 1858–1867. doi:10.1109/tnsre.2018.2864119.

Goodfellow, I, Bengio, Y, and Courville, A (2016) Deep learning (adaptive computation and machine learning series).

Goodfellow, IJ, Pouget-Abadie, J, Mirza, M, … Bengio, Y (2014) Generative Adversarial Networks. ArXiv. doi:10.48550/arxiv.1406.2661.

Gray, H, Guan, S, Loan, P, and Pandy, M (2018) Handbook of Human Motion., 101–115. doi:10.1007/978-3-319-14418-4_154.

Grisetti, G, Kummerle, R, Stachniss, C, and Burgard, W (2010) A Tutorial on Graph-Based SLAM. IEEE Intelligent Transportation Systems Magazine, 2(4), 31–43. doi:10.1109/mits.2010.939925.

Guan, S, Gray, HA, Keynejad, F, and Pandy, MG (2016) Mobile Biplane X-Ray Imaging System for Measuring 3D Dynamic Joint Motion During Overground Gait. IEEE Transactions on Medical Imaging, 35(1), 326–336. doi:10.1109/tmi.2015.2473168.

Halilaj, E, Rajagopal, A, Fiterau, M, Hicks, JL, Hastie, TJ, and Delp, SL (2018) Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of Biomechanics, 81, 1–11. doi:10.1016/j.jbiomech.2018.09.009.

Hamidon, N, Azaman, A, Mahmod, SR, and Abdullah, NHO (2022) Validity and Repeatability of Inertial Measurement Unit for Measuring Walking Gait Parameter of

Patients with Non-specific Low Back Pain. Malaysian Journal of Medicine and Health Sciences, 45–51. doi:10.47836/mjmhs.18.s6.9.

Han, K, Wang, Y, Chen, H, … Tao, D (2022) A Survey on Vision Transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 1–1. doi:10.1109/tpami.2022.3152247.

Hannink, J, Kautz, T, Pasluosta, CF, … Eskofier, BM (2016a) Stride Length Estimation with Deep Learning. ArXiv. doi:10.48550/arxiv.1609.03321.

Hannink, J, Kautz, T, Pasluosta, CF, … Eskofier, BM (2018) Mobile Stride Length Estimation With Deep Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics, 22(2), 354–362. doi:10.1109/jbhi.2017.2679486.

Hannink, J, Kautz, T, Pasluosta, CF, Gasmann, K-G, Klucken, J, and Eskofier, BM (2016b) Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics, 21(1), 85–93. doi:10.1109/jbhi.2016.2636456.

Hannink, J, Ollenschläger, M, Kluge, F, Roth, N, Klucken, J, and Eskofier, BM (2017) Benchmarking Foot Trajectory Estimation Methods for Mobile Gait Analysis. Sensors, 17(9), 1940. doi:10.3390/s17091940.

Hantouly, AT, Ahmed, AF, Alzobi, O, … Dosari, MAAA (2022) Mobile-bearing versus fixed-bearing total knee arthroplasty: a meta-analysis of randomized controlled trials. European Journal of Orthopaedic Surgery & Traumatology, 32(3), 481–495. doi:10.1007/s00590-021-02999-x.

Hatfield, GL, Hubley-Kozey, CL, Wilson, JLA, and Dunbar, MJ (2011) The Effect of Total Knee Arthroplasty on Knee Joint Kinematics and Kinetics During Gait. The Journal of Arthroplasty, 26(2), 309–318. doi:10.1016/j.arth.2010.03.021.

Hausdorff, JM (2005) Gait variability: methods, modeling and meaning. Journal of NeuroEngineering and Rehabilitation, 2(1), 19. doi:10.1186/1743-0003-2-19.

He, K, Zhang, X, Ren, S, and Sun, J (2015) Deep Residual Learning for Image Recognition. ArXiv. doi:10.48550/arxiv.1512.03385.

Hendry, D, Leadbetter, R, McKee, K, … Campbell, A (2020) An Exploration of Machine-Learning Estimation of Ground Reaction Force from Wearable Sensor Data. Sensors, 20(3), 740. doi:10.3390/s20030740.

Henriksen, M, Graven-Nielsen, T, Aaboe, J, Andriacchi, TP, and Bliddal, H (2010) Gait changes in patients with knee osteoarthritis are replicated by experimental knee pain. Arthritis Care & Research, 62(4), 501–509. doi:10.1002/acr.20033.

Hernandez, V, Dadkhah, D, Babakeshizadeh, V, and Kulić, D (2021) Lower body kinematics estimation from wearable sensors for walking and running: A deep learning approach. Gait & Posture, 83, 185–193. doi:10.1016/j.gaitpost.2020.10.026.

Hernandez, V, Suzuki, T, and Venture, G (2020) Convolutional and recurrent neural network for human activity recognition: Application on American sign language. PLoS ONE, 15(2), e0228869. doi:10.1371/journal.pone.0228869.

Hinton, G, Vinyals, O, and Dean, J (2015) Distilling the Knowledge in a Neural Network.

Hochreiter, S, and Schmidhuber, J (1997) Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hollman, JH, McDade, EM, and Petersen, RC (2011) Normative spatiotemporal gait parameters in older adults. Gait & Posture, 34(1), 111–118. doi:10.1016/j.gaitpost.2011.03.024.

Hootman, JM, Helmick, CG, Barbour, KE, Theis, KA, and Boring, MA (2016) Updated Projected Prevalence of Self-Reported Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation Among US Adults, 2015–2040. Arthritis & Rheumatology, 68(7), 1582–1587. doi:10.1002/art.39692.

Hossain, MSB, Dranetz, J, Choi, H, Guo, Z, and Guo, Z (2022) DeepBBWAE-Net: A CNN-RNN Based Deep SuperLearner for Estimating Lower Extremity Sagittal Plane Joint Kinematics Using Shoe-Mounted IMU Sensors in Daily Living. IEEE Journal of Biomedical and Health Informatics, 26(8), 3906–3917. doi:10.1109/jbhi.2022.3165383.

Hsu, Y-C, Shen, Y, Jin, H, and Kira, Z (2020) Generalized ODIN: Detecting Out-of-distribution Image without Learning from Out-of-distribution Data.

Hu, B, Dixon, PC, Jacobs, JV, Dennerlein, JT, and Schiffman, JM (2018) Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking. Journal of Biomechanics, 71, 37–42. doi:10.1016/j.jbiomech.2018.01.005.

Hume, DR, Kefala, V, Harris, MD, and Shelburne, KB (2018) Comparison of Marker-Based and Stereo Radiography Knee Kinematics in Activities of Daily Living. Annals of Biomedical Engineering, 46(11), 1806–1815. doi:10.1007/s10439-018-2068-9.

Imtiaz, U, Yamamura, K, Kong, W, … Takanishi, A (2014) Application of Wireless Inertial Measurement Units and EMG Sensors for studying Deglutition - Preliminary Results. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014, 5381–5384. doi:10.1109/embc.2014.6944842.

Inman, VT, Ralston, HJ, and Todd, F (1981) Human walking (W. & Wilkins, Ed.).

Ishii, Y, Terajima, K, Koga, Y, Takahashi, HE, Bechtold, JE, and Gustilo, RB (1998) Gait analysis after total knee arthroplasty. Comparison of posterior cruciate retention and substitution. Journal of Orthopaedic Science, 3(6), 310–317. doi:10.1007/s007760050058.

Johnson, WR, Mian, A, Lloyd, DG, and Alderson, JA (2019) On-field player workload exposure and knee injury risk monitoring via deep learning. Journal of Biomechanics, 93, 185–193. doi:10.1016/j.jbiomech.2019.07.002.

Johnson, WR, Mian, A, Robinson, MA, Verheul, J, Lloyd, DG, and Alderson, J (2020) Multidimensional ground reaction forces and moments from wearable sensor accelerations via deep learning. IEEE Transactions on Biomedical Engineering, PP(99), 1–1. doi:10.1109/tbme.2020.3006158.

Karatsidis, A, Jung, M, Schepers, HM, … Andersen, MS (2018) Predicting kinetics using musculoskeletal modeling and inertial motion capture. ArXiv.

Karatsidis, A, Jung, M, Schepers, HM, … Andersen, MS (2019) Musculoskeletal model-based inverse dynamic analysis under ambulatory conditions using inertial motion capture. Medical Engineering & Physics, 65, 68–77. doi:10.1016/j.medengphy.2018.12.021.

Karim, F, Majumdar, S, Darabi, H, and Harford, S (2018) Multivariate LSTM-FCNs for Time Series Classification. ArXiv. doi:10.1016/j.neunet.2019.04.014.

Kaufman, KR, Hughes, C, Morrey, BF, Morrey, M, and An, K-N (2001) Gait characteristics of patients with knee osteoarthritis. Journal of Biomechanics, 34(7), 907–915. doi:10.1016/s0021-9290(01)00036-7.

Kautz, T, Groh, BH, Hannink, J, Jensen, U, Strubberg, H, and Eskofier, BM (2017) Activity recognition in beach volleyball using a Deep Convolutional Neural Network. Data Mining and Knowledge Discovery, 31(6), 1678–1705. doi:10.1007/s10618-017-0495-0.

Khera, P, and Kumar, N (2020) Role of machine learning in gait analysis: a review. Journal of Medical Engineering & Technology, 44(8), 441–467. doi:10.1080/03091902.2020.1822940.

Kim, J-H, Zhang, Y, Han, K, Wen, Z, Choi, M, and Liu, Z (2021) Representation learning of resting state fMRI with variational autoencoder. NeuroImage, 241, 118423. doi:10.1016/j.neuroimage.2021.118423.

Kingma, DP, and Ba, J (2014) Adam: A Method for Stochastic Optimization. ArXiv. doi:10.48550/arxiv.1412.6980.

Kingma, DP, and Welling, M (2013) Auto-Encoding Variational Bayes. ArXiv. doi:10.48550/arxiv.1312.6114.

Kiss, RM (2011) Effect of severity of knee osteoarthritis on the variability of gait parameters. Journal of Electromyography and Kinesiology, 21(5), 695–703. doi:10.1016/j.jelekin.2011.07.011.

Kiss, RM, Bejek, Z, and Szendrői, M (2012) Variability of gait parameters in patients with total knee arthroplasty. Knee Surgery, Sports Traumatology, Arthroscopy, 20(7), 1252–1260. doi:10.1007/s00167-012-1965-y.

Klette, R, and Tee, G (2008) Understanding Human Motion: A Historic Review. doi:https://doi.org/10.1007/978-1-4020-6693-1_1.

Kluge, F, Gaßner, H, Hannink, J, Pasluosta, C, Klucken, J, and Eskofier, BM (2017) Towards Mobile Gait Analysis: Concurrent Validity and Test-Retest Reliability of an Inertial Measurement System for the Assessment of Spatio-Temporal Gait Parameters. Sensors, 17(7), 1522. doi:10.3390/s17071522.

Komaris, D-S, Govind, C, Murphy, AJ, … Riches, P (2021) Implant design affects walking and stair navigation after total knee arthroplasty: a double-blinded randomised controlled trial. Journal of Orthopaedic Surgery and Research, 16(1), 177. doi:10.1186/s13018-021-02311-x.

Komaris, D-S, Pérez-Valero, E, Jordan, L, … Tedesco, S (2019) Predicting Three-Dimensional Ground Reaction Forces in Running by Using Artificial Neural Networks and Lower Body Kinematics. IEEE Access, 7, 156779–156786. doi:10.1109/access.2019.2949699.

Konrath, J, Karatsidis, A, Schepers, H, Bellusci, G, Zee, M de, and Andersen, M (2019) Estimation of the Knee Adduction Moment and Joint Contact Force during Daily Living Activities Using Inertial Motion Capture. Sensors, 19(7), 1681. doi:10.3390/s19071681.

Krishnapuram, B, Shah, M, Smola, A, … Guestrin, C (2016) XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. doi:10.1145/2939672.2939785.

Kurtz, SM, Higgs, GB, Chen, Z, … Mont, MA (2022) Patient Perceptions of Wearable and Smartphone Technologies for Remote Outcome Monitoring in Patients Who Have Hip Osteoarthritis or Arthroplasties. The Journal of Arthroplasty, 37(7), S488-S492.e2. doi:10.1016/j.arth.2022.02.026.

Lee, A, Park, J, and Lee, S (2015) Gait analysis of elderly women after total knee arthroplasty. Journal of Physical Therapy Science, 27(3), 591–595. doi:10.1589/jpts.27.591.

Levinger, P, Menz, HB, Morrow, AD, … Bergman, NB (2011) Knee biomechanics early after knee replacement surgery predict abnormal gait patterns 12 months postoperatively. Journal of Orthopaedic Research : Official Publication of the Orthopaedic Research Society, 30(3), 371–6. doi:10.1002/jor.21545.

Levinger, P, Menz, HB, Morrow, AD, Feller, JA, Bartlett, JR, and Bergman, NR (2013) Lower Limb Biomechanics in Individuals With Knee Osteoarthritis Before and After Total Knee Arthroplasty Surgery. The Journal of Arthroplasty, 28(6), 994–999. doi:10.1016/j.arth.2012.10.018.

Liang, S, Li, Y, and Srikant, R (2017) Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks. ArXiv. doi:10.48550/arxiv.1706.02690.

Lim, Kim, and Park (2019) Prediction of Lower Limb Kinetics and Kinematics during Walking by a Single IMU on the Lower Back Using Machine Learning. Sensors, 20(1), 130. doi:10.3390/s20010130.

Liu, D-X, Du, W, Wu, X, Wang, C, and Qiao, Y (2016) Deep Rehabilitation Gait Learning for Modeling Knee Joints of Lower-Limb Exoskeleton. 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), 1058–1063. doi:10.1109/robio.2016.7866465.

Lopez-Nava, IH, and Munoz-Melendez, A (2016) Wearable Inertial Sensors for Human Motion Analysis: A Review. IEEE Sensors Journal, 16(22), 7821–7834. doi:10.1109/jsen.2016.2609392.

McCabe, MV (2020) Utilizing Neural Networks and Wearables to Quantify Hip Joint Angles and Moments During Walking and Stair Ascent, https://digitalcommons.dartmouth.edu/engs88.

Mentiplay, BF, Perraton, LG, Bower, KJ, … Clark, RA (2015) Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. Journal of Biomechanics, 48(10), 2166–70. doi:10.1016/j.jbiomech.2015.05.021.

Meyer, CAG, Corten, K, Fieuws, S, … Desloovere, K (2016) Evaluation of stair motion contributes to new insights into hip osteoarthritis-related motion pathomechanics. Journal of Orthopaedic Research, 34(2), 187–196. doi:10.1002/jor.22990.

Migliorini, F, Maffulli, N, Cuozzo, F, Pilone, M, Elsner, K, and Eschweiler, J (2022) No difference between mobile and fixed bearing in primary total knee arthroplasty: a meta-analysis. Knee Surgery, Sports Traumatology, Arthroscopy, 30(9), 3138–3154. doi:10.1007/s00167-022-07065-5.

Moissenet, F, Leboeuf, F, and Armand, S (2019) Lower limb sagittal gait kinematics can be predicted based on walking speed, gender, age and BMI. Scientific Reports, 9(1), 9510. doi:10.1038/s41598-019-45397-4.

Morris, JRW (1973) Accelerometry—A technique for the measurement of human body movements. Journal of Biomechanics, 6(6), 729–736. doi:10.1016/0021-9290(73)90029-8.

Mündermann, L, Corazza, S, and Andriacchi, TP (2006) The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. Journal of NeuroEngineering and Rehabilitation, 3(1), 6. doi:10.1186/1743-0003-3-6.

Mundt, M, Johnson, WR, Potthast, W, Markert, B, Mian, A, and Alderson, J (2021) A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units. Sensors, 21(13), 4535. doi:10.3390/s21134535.

Mundt, M, Koeppe, A, Bamer, F, David, S, and Markert, B (2020a) Artificial Neural Networks in Motion Analysis—Applications of Unsupervised and Heuristic Feature Selection Techniques. Sensors, 20(16), 4581. doi:10.3390/s20164581.

Mundt, M, Koeppe, A, David, S, … Markert, B (2020b) Estimation of Gait Mechanics Based on Simulated and Measured IMU Data Using an Artificial Neural Network. Frontiers in Bioengineering and Biotechnology, 8, 41. doi:10.3389/fbioe.2020.00041.

Mundt, M, Thomsen, W, Witter, T, … Markert, B (2019) Prediction of lower limb joint angles and moments during gait using artificial neural networks. Medical & Biological Engineering & Computing, 58(1), 211–225. doi:10.1007/s11517-019-02061-3.

Murdoch, B (2021) Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Medical Ethics, 22(1), 122. doi:10.1186/s12910-021-00687-3.

Myers, CA, Laz, PJ, Shelburne, KB, and Davidson, BS (2015) A Probabilistic Approach to Quantify the Impact of Uncertainty Propagation in Musculoskeletal Simulations. Annals of Biomedical Engineering, 43(5), 1098–1111. doi:10.1007/s10439-014-1181-7.

Myers, CA, Laz, PJ, Shelburne, KB, … Rullkoetter, PJ (2018) The impact of hip implant alignment on muscle and joint loading during dynamic activities. Clinical Biomechanics, 53, 93–100. doi:10.1016/j.clinbiomech.2018.02.010.

Nahavandi, D, Alizadehsani, R, Khosravi, A, and Acharya, UR (2021) Application of artificial intelligence in wearable devices: Opportunities and Challenges. Computer Methods and Programs in Biomedicine, 213, 106541. doi:10.1016/j.cmpb.2021.106541.

Nigg, BM, and Herzog, W (2007) Biomechanics of the Musculo-skeletal System.

Norvig, P, and Russell, SJ (2020) Artificial Intelligence: A Modern Approach.

Nweke, HF, Teh, YW, Al-garadi, MA, and Alo, UR (2018) Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges. Expert Systems with Applications, 105, 233–261. doi:10.1016/j.eswa.2018.03.056.

O'Donovan, KJ, Kamnik, R, O'Keeffe, DT, and Lyons, GM (2007) An inertial and magnetic sensor based technique for joint angle measurement. Journal of Biomechanics, 40(12), 2604–2611. doi:10.1016/j.jbiomech.2006.12.010.

Oguiza, I (2022) tsai - A state-of-the-art deep learning library for time series and sequential data. Retrieved from https://github.com/timeseriesAI/tsai

Oh, J, Kuenze, C, Jacopetti, M, Signorile, JF, and Eltoukhy, M (2018) Validity of the Microsoft KinectTM in assessing spatiotemporal and lower extremity kinematics during stair ascent and descent in healthy young individuals. Medical Engineering & Physics, 60, 70–76. doi:10.1016/j.medengphy.2018.07.011.

OHTAKI, Y, SAGAWA, K, and INOOKA, H (2001) A Method for Gait Analysis in a Daily Living Environment by Body-Mounted Instruments. JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing, 44(4), 1125–1132. doi:10.1299/jsmec.44.1125.

Ordóñez, FJ, and Roggen, D (2016) Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors, 16(1), 115. doi:10.3390/s16010115.

Outerleys, JB, Dunbar, MJ, Richardson, G, Hubley-Kozey, CL, and Wilson, JLA (2021) Quantifying Achievable Levels of Improvement in Knee Joint Biomechanics During Gait After Total Knee Arthroplasty Relative to Osteoarthritis Severity. Journal of Applied Biomechanics, 37(2), 130–138. doi:10.1123/jab.2020-0051.

Panero, E, Digo, E, Agostini, V, and Gastaldi, L (2018) Comparison of Different Motion Capture Setups for Gait Analysis : Validation of spatio-temporal parameters estimation. 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 1–6. doi:10.1109/memea.2018.8438653.

Patterson, MR, Johnston, W, O'Mahony, N, … O'Mahony, S (2016) Validation of Temporal Gait Metrics from Three IMU Locations to the Gold Standard Force Plate. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, 667–671. doi:10.1109/embc.2016.7590790.

Paulo, J, Peixoto, P, and Amorim, P (2019) Trajectory-based gait pattern shift detection for assistive robotics applications. Intelligent Service Robotics, 12(3), 255–264. doi:10.1007/s11370-019-00280-z.

Picerno, P (2017) 25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches. Gait & Posture, 51, 239–246. doi:10.1016/j.gaitpost.2016.11.008.

Picerno, P, Cereatti, A, and Cappozzo, A (2008) Joint kinematics estimate using wearable inertial and magnetic sensing modules. Gait & Posture, 28(4), 588–595. doi:10.1016/j.gaitpost.2008.04.003.

Piche, E, Guilbot, M, Chorin, F, Guerin, O, Zory, R, and Gerus, P (2022) Validity and repeatability of a new inertial measurement unit system for gait analysis on kinematic parameters: Comparison with an optoelectronic system. Measurement, 198, 111442. doi:10.1016/j.measurement.2022.111442.

Ramkumar, PN, Haeberle, HS, Ramanathan, D, … Patterson, BM (2019) Remote Patient Monitoring Using Mobile Health for Total Knee Arthroplasty: Validation of a Wearable and Machine Learning–Based Surveillance Platform. The Journal of Arthroplasty, 34(10), 2253–2259. doi:10.1016/j.arth.2019.05.021.

Rampp, A, Barth, J, Schuelein, S, Gassmann, K-G, Klucken, J, and Eskofier, BM (2014) Inertial Sensor-Based Stride Parameter Calculation From Gait Sequences in Geriatric Patients. IEEE Transactions on Biomedical Engineering, 62(4), 1089–1097. doi:10.1109/tbme.2014.2368211.

Rao, L, Taylor, WR, Horn, N, List, R, Preiss, S, and Schütz, P (2022) Can tibio-femoral kinematic and kinetic parameters reveal poor functionality and underlying deficits after total knee replacement? A systematic review. The Knee, 34, 62–75. doi:10.1016/j.knee.2021.11.002.

Rapp, E, Shin, S, Thomsen, W, Ferber, R, and Halilaj, E (2021) Estimation of kinematics from inertial measurement units using a combined deep learning and optimization framework. Journal of Biomechanics, 116, 110229. doi:10.1016/j.jbiomech.2021.110229.

Raschka, S (2018) Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. ArXiv.

Reid, SM, Graham, RB, and Costigan, PA (2010) Differentiation of young and older adult stair climbing gait using principal component analysis. Gait & Posture, 31(2), 197–203. doi:10.1016/j.gaitpost.2009.10.005.

Renani, MS, Eustace, AM, Myers, CA, and Clary, CW (2021) The Use of Synthetic IMU Signals in the Training of Deep Learning Models Significantly Improves the Accuracy of Joint Kinematic Predictions. Sensors (Basel, Switzerland), 21(17), 5876. doi:10.3390/s21175876.

Renani, MS, Myers, CA, Zandie, R, Mahoor, MH, Davidson, BS, and Clary, CW (2020) Deep Learning in Gait Parameter Prediction for OA and TKA Patients Wearing IMU Sensors. Sensors, 20(19), 5553. doi:10.3390/s20195553.

Rezaei, M, and Shahidi, M (2020) Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. Intelligence-Based Medicine, 3, 100005. doi:10.1016/j.ibmed.2020.100005.

Rezende, DJ, and Mohamed, S (2015) Variational Inference with Normalizing Flows. ArXiv. doi:10.48550/arxiv.1505.05770.

RITTMAN, N, KETTELKAMP, DB, PRYOR, P, SCHWARTZKOPF, GL, and HILLBERRY, B (1981) Analysis of Patterns of Knee Motion Walking for Four Types of Total Knee Implants. Clinical Orthopaedics and Related Research, 155(NA;), 111–117. doi:10.1097/00003086-198103000-00021.

Ro, DH, Kang, T, Han, D hwan, Lee, DY, Han, H-S, and Lee, MC (2020) Quantitative evaluation of gait features after total knee arthroplasty: Comparison with age and sex-matched controls. Gait & Posture, 75, 78–84. doi:10.1016/j.gaitpost.2019.09.026.

Ro, DH, Lee, J, Lee, J, Park, JY, Han, HS, and Lee, MC (2019) Effects of Knee Osteoarthritis on Hip and Ankle Gait Mechanics. Advances in Orthopedics, 2019(1), 1–6. doi:10.1155/2019/9757369.

Robbins, SM, Wilson, JLA, Rutherford, DJ, and Hubley-Kozey, CL (2013) Reliability of principal components and discrete parameters of knee angle and moment gait waveforms in individuals with moderate knee osteoarthritis. Gait & Posture, 38(3), 421–427. doi:10.1016/j.gaitpost.2013.01.001.

Romijnders, R, Warmerdam, E, Hansen, C, Schmidt, G, and Maetzler, W (2022) A Deep Learning Approach for Gait Event Detection from a Single Shank-Worn IMU: Validation in Healthy and Neurological Cohorts. Sensors (Basel, Switzerland), 22(10), 3859. doi:10.3390/s22103859.

Rosenblatt, F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408. doi:10.1037/h0042519.

Rosso, V, Agostini, V, Takeda, R, Tadano, S, and Gastaldi, L (2019) Influence of BMI on Gait Characteristics of Young Adults: 3D Evaluation Using Inertial Sensors. Sensors (Basel, Switzerland), 19(19), 4221. doi:10.3390/s19194221.

Ruan, B-K, Shuai, H-H, and Cheng, W-H (2022) Vision Transformers: State of the Art and Research Challenges. ArXiv. doi:10.48550/arxiv.2207.03041.

Ruiz, N, Schulter, S, and Chandraker, M (2018) Learning To Simulate. ArXiv. doi:10.48550/arxiv.1810.02513.

Ryan, J, Mora, JP, Scuderi, GR, and Jr., AJT (2021) Total Knee Arthroplasty Design and Kinematics: Past, Present, and Future. Journal of Long-Term Effects of Medical Implants, 31(3), 1–14. doi:10.1615/jlongtermeffmedimplants.2021038212.

Samuel, AL (1959) Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3), 210–229. doi:10.1147/rd.33.0210.

Sartori, M, Lloyd, DG, and Farina, D (2016a) Corrections to “Neural Data-Driven Musculoskeletal Modeling for Personalized Neurorehabilitation Technologies” [May 16 879-893]. IEEE Transactions on Biomedical Engineering, 63(6), 1341–1341. doi:10.1109/tbme.2016.2563138.

Sartori, M, Llyod, DG, and Farina, D (2016b) Neural Data-Driven Musculoskeletal Modeling for Personalized Neurorehabilitation Technologies. IEEE Transactions on Biomedical Engineering, 63(5), 879–893. doi:10.1109/tbme.2016.2538296.

Schütz, P, Postolka, B, Gerber, H, Ferguson, SJ, Taylor, WR, and List, R (2019) Knee implant kinematics are task-dependent. Journal of the Royal Society Interface, 16(151), 20180678. doi:10.1098/rsif.2018.0678.

Seel, T, Raisch, J, and Schauer, T (2014) IMU-Based Joint Angle Measurement for Gait Analysis. Sensors, 14(4), 6891–6909. doi:10.3390/s140406891.

Siami-Namini, S, Tavakoli, N, and Namin, AS (2019) The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data), 00, 3285–3292. doi:10.1109/bigdata47090.2019.9005997.

Siddhad, G, Gupta, A, Dogra, DP, and Roy, PP (2022) Efficacy of Transformer Networks for Classification of Raw EEG Data. ArXiv. doi:10.48550/arxiv.2202.05170.

Singh, JP, Jain, S, Arora, S, and Singh, UP (2018) Vision-Based Gait Recognition: A Survey. IEEE Access, 6, 70497–70527. doi:10.1109/access.2018.2879896.

Skals, S, Rasmussen, KP, Bendtsen, KM, Yang, J, and Andersen, MS (2017) A musculoskeletal model driven by dual Microsoft Kinect Sensor data. Multibody System Dynamics, 41(4), 297–316. doi:10.1007/s11044-017-9573-8.

Skipper, A, Yang;Jian, Zee;Mark, de, Zhou;Lelai, Bai;Shaoping, and Rasmussen;John (2013) Full-body musculoskeletal modeling using dual microsoft kinect sensors and the anybody modeling system. International Symposium on Computer Simulation in Biomechanics.

Snell, DL, Hipango, J, Sinnott, KA, … Hooper, G (2018) Rehabilitation after total joint replacement: a scoping study. Disability and Rehabilitation, 40(14), 1718–1731. doi:10.1080/09638288.2017.1300947.

Sohl-Dickstein, J, Weiss, EA, Maheswaranathan, N, and Ganguli, S (2015) Deep Unsupervised Learning using Nonequilibrium Thermodynamics. ArXiv. doi:10.48550/arxiv.1503.03585.

Soltani, A, Dejnabadi, H, Savary, M, and Aminian, K (2018) Real-World Gait Speed Estimation Using Wrist Sensor: A Personalized Approach. IEEE Journal of Biomedical and Health Informatics, 24(3), 658–668. doi:10.1109/jbhi.2019.2914940.

Sparkes, V, Whatling, GM, Biggs, P, … Holt, C (2019) Comparison of gait, functional activities, and patient-reported outcome measures in patients with knee osteoarthritis and healthy adults using 3D motion analysis and activity monitoring: an exploratory case-control analysis. Orthopedic Research and Reviews, 11, 129–140. doi:10.2147/orr.s199107.

SRJ, and Everitt, BS (1999) The Cambridge Dictionary of Statistics. Journal of the American Statistical Association, 94(446), 657. doi:10.2307/2670205.

Standifird, TW (2015) Lower Extremity Biomechanics during Stair Ascent in Healthy and Total Knee Replacement Older Adults.

Stetter, BJ, Krafft, FC, Ringhof, S, Stein, T, and Sell, S (2020) A Machine Learning and Wearable Sensor Based Approach to Estimate External Knee Flexion and Adduction Moments During Various Locomotion Tasks. Frontiers in Bioengineering and Biotechnology, 8, 9. doi:10.3389/fbioe.2020.00009.

Sumner, B, McCamley, J, Jacofsky, D, and Jacofsky, M (2019) Comparison of Knee Kinematics and Kinetics during Stair Descent in Single- and Multi-Radius Total Knee Arthroplasty. The Journal of Knee Surgery, 33(10), 1020–1028. doi:10.1055/s-0039-1692652.

Sun, J, Xie, J, and Zhou, H (2021) EEG Classification with Transformer-Based Models. 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), 00, 92–93. doi:10.1109/lifetech52111.2021.9391844.

Szopa, A, Domagalska-Szopa, M, Siwiec, A, and Kwiecień-Czerwieniec, I (2020) Canonical correlation between body-posture deviations and gait disorders in children with cerebral palsy. PLOS ONE, 15(6), e0234654. doi:10.1371/journal.pone.0234654.

Tan, C, Sun, F, Kong, T, Zhang, W, Yang, C, and Liu, C (2018) A Survey on Deep Transfer Learning. ArXiv. doi:10.48550/arxiv.1808.01974.

Tan, J-S, Tippaya, S, Binnie, T, … Campbell, A (2022) Predicting Knee Joint Kinematics from Wearable Sensor Data in People with Knee Osteoarthritis and Clinical Considerations for Future Machine Learning Models. Sensors (Basel, Switzerland), 22(2), 446. doi:10.3390/s22020446.

Tazawa, M, Sohmiya, M, Wada, N, Defi, IR, and Shirakura, K (2014) Toe-out angle changes after total knee arthroplasty in patients with varus knee osteoarthritis. Knee Surgery, Sports Traumatology, Arthroscopy, 22(12), 3168–3173. doi:10.1007/s00167-014-2841-8.

Tobore, I, Li, J, Yuhang, L, … Wang, L (2019) Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. JMIR MHealth and UHealth, 7(8), e11966. doi:10.2196/11966.

Tran, L, and Choi, D (2020) Data Augmentation for Inertial Sensor-Based Gait Deep Neural Network. IEEE Access, 8, 12364–12378. doi:10.1109/access.2020.2966142.

Trawiński, B, Smętek, M, Telec, Z, and Lasota, T (2012) Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. International Journal of Applied Mathematics and Computer Science, 22(4), 867–881. doi:10.2478/v10006-012-0064-z.

Trinler, UK, Baty, F, Mündermann, A, … Wegener, R (2016) Stair dimension affects knee kinematics and kinetics in patients with good outcome after TKA similarly as in healthy subjects: STAIR CLIMBING IN TKA PATIENTS. Journal of Orthopaedic Research, 34(10), 1753–1761. doi:10.1002/jor.23181.

Trojaniello, D, Cereatti, A, Pelosin, E, … Croce, UD (2014) Estimation of step-by-step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic, parkinsonian and choreic gait. Journal of NeuroEngineering and Rehabilitation, 11(1), 152. doi:10.1186/1743-0003-11-152.

Um, TT, Pfister, FMJ, Pichler, D, … Kulić, D (2017) Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. ArXiv, 216–220. doi:10.1145/3136755.3136817.

Usami, T, Nishida, K, Iguchi, H, … Kuroyanagi, G (2022) Evaluation of lower extremity gait analysis using Kinect V2® tracking system. SICOT-J, 8, 27. doi:10.1051/sicotj/2022027.

Vargas-Valencia, LS, Elias, A, Rocon, E, Bastos-Filho, T, and Frizera, A (2016) An IMU-to-Body Alignment Method Applied to Human Gait Analysis. Sensors, 16(12), 2090. doi:10.3390/s16122090.

Vaswani, A, Shazeer, N, Parmar, N, … Polosukhin, I (2017) Attention Is All You Need. ArXiv.

Veilleux, L-N, Raison, M, Rauch, F, Robert, M, and Ballaz, L (2016) Agreement of spatio-temporal gait parameters between a vertical ground reaction force decomposition algorithm and a motion capture system. Gait & Posture, 43, 257–264. doi:10.1016/j.gaitpost.2015.10.007.

Vydhyanathan, A, Bellusci, G, Luinge, H, and Slycke, and P (2015) The Next Generation Xsens Motion Trackers for Industrial Applications. Xsens: Enschede. Retrieved from http://www.souvr.com/UploadFiles/Soft_UploadSoft/201405/MTi%20%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf

Wang, J, Lan, C, Liu, C, Ouyang, Y, Zeng, W, and Qin, T (2021) Generalizing to Unseen Domains: A Survey on Domain Generalization. ArXiv.

Wang, Z, Yan, W, and Oates, T (2017) Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. 2017 International Joint Conference on Neural Networks (IJCNN), 1578–1585. doi:10.1109/ijcnn.2017.7966039.

Washabaugh, EP, Kalyanaraman, T, Adamczyk, PG, Claflin, ES, and Krishnan, C (2017) Validity and repeatability of inertial measurement units for measuring gait parameters. Gait & Posture, 55, 87–93. doi:10.1016/j.gaitpost.2017.04.013.

Weiss, A, Brozgol, M, Dorfman, M, … Hausdorff, JM (2013) Does the Evaluation of Gait Quality During Daily Life Provide Insight Into Fall Risk? A Novel Approach Using 3-Day Accelerometer Recordings. Neurorehabilitation and Neural Repair, 27(8), 742–752. doi:10.1177/1545968313491004.

Wen, Q, Sun, L, Song, X, Gao, J, Wang, X, and Xu, H (2020) Time Series Data Augmentation for Deep Learning: A Survey. ArXiv.

Weygers, I, Kok, M, Konings, M, Hallez, H, Vroey, HD, and Claeys, K (2020) Inertial Sensor-Based Lower Limb Joint Kinematics: A Methodological Systematic Review. Sensors, 20(3), 673. doi:10.3390/s20030673.

Whitchelo, T, McClelland, JA, and Webster, KE (2013) Factors associated with stair climbing ability in patients with knee osteoarthritis and knee arthroplasty: a systematic review. Disability and Rehabilitation, 36(13), 1051–1060. doi:10.3109/09638288.2013.829526.

Willemsen, AThM, Alsté, JA van, and Boom, HBK (1990) Real-time gait assessment utilizing a new way of accelerometry. Journal of Biomechanics, 23(8), 859–863. doi:10.1016/0021-9290(90)90033-y.

Wilson, JLA, Dunbar, MJ, and Hubley-Kozey, CL (2015) Knee Joint Biomechanics and Neuromuscular Control During Gait Before and After Total Knee Arthroplasty are Sex-specific. The Journal of Arthroplasty, 30(1), 118–125. doi:10.1016/j.arth.2014.07.028.

Winter, DA (1983) Biomechanical Motor Patterns in Normal Walking. Journal of Motor Behavior, 15(4), 302–330. doi:10.1080/00222895.1983.10735302.

Witjes, S, Gouttebarge, V, Kuijer, PPFM, Geenen, RCI van, Poolman, RW, and Kerkhoffs, GMMJ (2016) Return to Sports and Physical Activity After Total and Unicondylar Knee Arthroplasty: A Systematic Review and Meta-Analysis. Sports Medicine, 46(2), 269–292. doi:10.1007/s40279-015-0421-9.

Wouda, FJ, Giuberti, M, Bellusci, G, … Veltink, PH (2018) Estimation of Vertical Ground Reaction Forces and Sagittal Knee Kinematics During Running Using Three Inertial Sensors. Frontiers in Physiology, 9, 218. doi:10.3389/fphys.2018.00218.

Wu, N, Green, B, Ben, X, and O'Banion, S (2020) Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. ArXiv.

Wu, X, Chu, L, Xiao, L, … Liu, Y (2017) Early Spatiotemporal Patterns and Knee Kinematics during Level Walking in Individuals following Total Knee Arthroplasty. Journal of Healthcare Engineering, 2017, 1–5. doi:10.1155/2017/7056469.

Xian, Y, Lampert, CH, Schiele, B, and Akata, Z (2017) Zero-Shot Learning -- A Comprehensive Evaluation of the Good, the Bad and the Ugly. ArXiv. doi:10.48550/arxiv.1707.00600.

Yamamoto, M, Shimatani, K, Hasegawa, M, Kurita, Y, Ishige, Y, and Takemura, H (2021) Accuracy of Temporo-Spatial and Lower Limb Joint Kinematics Parameters Using OpenPose for Various Gait Patterns With Orthosis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 2666–2675. doi:10.1109/tnsre.2021.3135879.

Yocum, D, Reinbolt, J, Weinhandl, JT, … Zhang, S (2021) Principal Component Analysis of Knee Joint Differences Between Bilateral and Unilateral Total Knee Replacement Patients During Level Walking. Journal of Biomechanical Engineering, 143(11). doi:10.1115/1.4051524.

Yoo, TK, Kim, SK, Choi, SB, Kim, DY, and Kim, DW (2013) Interpretation of Movement during Stair Ascent for Predicting Severity and Prognosis of Knee Osteoarthritis in Elderly Women Using Support Vector Machine. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, 192–196. doi:10.1109/embc.2013.6609470.

Young, AD, Ling, MJ, and Arvind, DK (2014) IMUSim: A simulation environment for inertial sensing algorithm design and evaluation., IEEE. Retrieved from https://ieeexplore.ieee.org/document/5779038

Young-Shand, KL, Dunbar, MJ, and Wilson, JLA (2020) Individual Gait Features Are Associated with Clinical Improvement After Total Knee Arthroplasty. JBJS Open Access, 5(2), e0038–e0038. doi:10.2106/jbjs.oa.19.00038.

Zaroug, A, Lai, DTH, Mudie, K, and Begg, R (2020) Lower Limb Kinematics Trajectory Prediction Using Long Short-Term Memory Neural Networks. Frontiers in Bioengineering and Biotechnology, 8, 362. doi:10.3389/fbioe.2020.00362.

Zeni, JA, and Higginson, JS (2009) Differences in gait parameters between healthy subjects and persons with moderate and severe knee osteoarthritis: A result of altered walking speed? Clinical Biomechanics, 24(4), 372–378. doi:10.1016/j.clinbiomech.2009.02.001.

Zerveas, G, Jayaraman, S, Patel, D, Bhamidipaty, A, and Eickhoff, C (2020) A Transformer-based Framework for Multivariate Time Series Representation Learning. ArXiv.

Zheng, X, Wang, M, and Ordieres-Meré, J (2018) Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0. Sensors, 18(7), 2146. doi:10.3390/s18072146.

Zhou, H, Zhang, S, Peng, J, … Zhang, W (2020a) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. ArXiv.

Zhou, K, Liu, Z, Qiao, Y, Xiang, T, and Loy, CC (2022) Domain Generalization: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, PP(99), 1–20. doi:10.1109/tpami.2022.3195549.

Zhou, Y, Romijnders, R, Hansen, C, … Lamoth, CJC (2020b) The detection of age groups by dynamic gait outcomes using machine learning approaches. Scientific Reports, 10(1), 4426. doi:10.1038/s41598-020-61423-2.

Zhuang, F, Qi, Z, Duan, K, … He, Q (2019) A Comprehensive Survey on Transfer Learning. ArXiv. doi:10.48550/arxiv.1911.02685.

Zijlstra, W, and Hof, AL (2003) Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait & Posture, 18(2), 1–10. doi:10.1016/s0966-6362(02)00190-x.

Zimmermann, T, Taetz, B, and Bleser, G (2018) IMU-to-Segment Assignment and Orientation Alignment for the Lower Body Using Deep Learning. Sensors, 18(1), 302. doi:10.3390/s18010302.

Zizzo, G, and Ren, L (2017) Position Tracking During Human Walking Using an Integrated Wearable Sensing System. Sensors, 17(12), 2866. doi:10.3390/s17122866.

Zoph, B, Yuret, D, May, J, and Knight, K (2016) Transfer Learning for Low-Resource Neural Machine Translation. ArXiv. doi:10.48550/arxiv.1604.02201.

Zrenner, M, Gradl, S, Jensen, U, Ullrich, M, and Eskofier, BM (2018) Comparison of Different Algorithms for Calculating Velocity and Stride Length in Running Using Inertial Measurement Units. Sensors, 18(12), 4194. doi:10.3390/s18124194.

Żuk, M, and Pezowicz, C (2015) Kinematic Analysis of a Six-Degrees-of-Freedom Model Based on ISB Recommendation: A Repeatability Analysis and Comparison with Conventional Gait Model. Applied Bionics and Biomechanics, 2015, 503713. doi:10.1155/2015/503713.
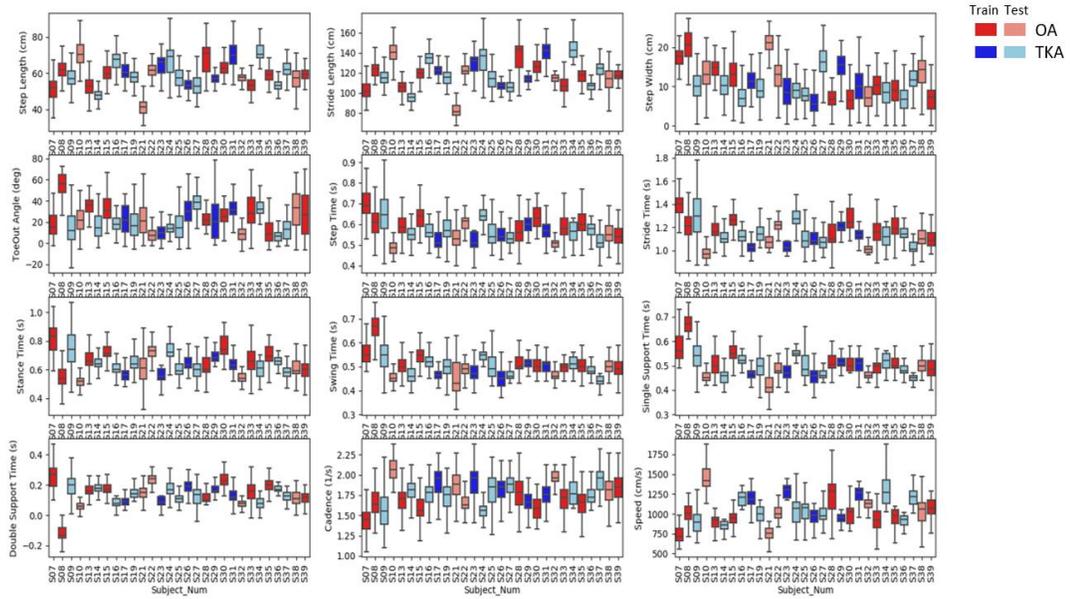
Figure A1: Boxplot associated with STGPs of each subject grouped by their knee status (OA and TKA) for training and test set.
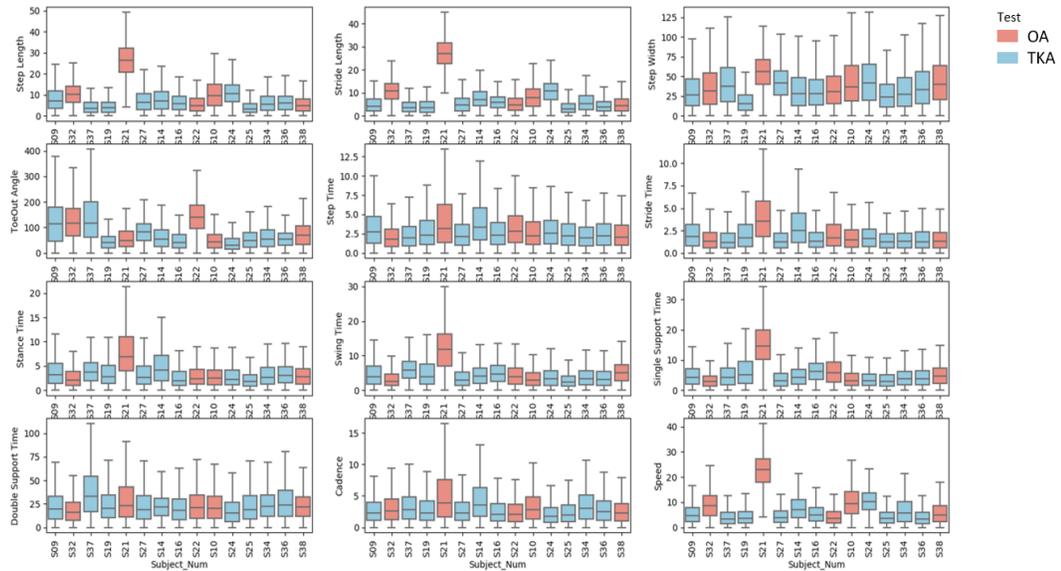
Figure A2: Boxplot associated with NAPEs of each test subjects grouped by their knee status (OA and TKA) for test set.

Table A1: Jensen–Shannon (JS) divergence (relative entropy) between each test subjects and training on fold with S19, S21, and S27 subjects.

	JS(Training‖S19)	JS(Training‖S21)	JS(Training‖S27)
Step Length	0.18	**4.67**	0.38
Stride Length	0.17	**7.01**	0.59
Step Width	0.10	**5.96**	0.67
Toe Out Angle	0.11	0.02	0.63
Step Time	0.01	0.26	0.80
Stride Time	0.02	0.31	1.26
Stance Time	0.01	0.14	0.63
Swing Time	0.02	0.33	0.55
Single Support Time	0.03	0.83	0.45
Double Support Time	0.08	0.04	0.29
Cadence	0.21	0.37	0.52
Speed	0.07	0.50	0.29

Table A2: Homogeneous subsets based on Freidman ranking and asymptotic significances (0.05) for a) OA, b) TKA cohort, c) slow pace, d) normal, and e) fast pace.

(a) OA	Subsets					
Sensors	**1**	**2**	**3**	**4**	**5**	**6**
T	7.333					
F S		7.620				
F T		7.695	7.695			
S		7.776	7.776			
F		7.779	7.779			
F P			7.803			
F P T			7.809			
F S T			7.866			
F P S			7.985	7.985		
F P S T				8.115		
P T				8.144		
S T				8.164		
P S					8.445	
P S T					8.590	
P						8.876
Test Statistic	.	9.049	9.052	9.927	6.536	.
Adjusted Sig. (2-sided test)	.	.103	.330	.070	.077	.

167

(b) TKA Sensors	Subset 1	2	3	4	5	6	7
F T	7.573						
F S	7.723	7.723					
F P T	7.807	7.807					
F P S T		7.822	7.822				
F S T		7.854	7.854				
S T		7.872	7.872				
S		7.878	7.878	7.878			
F		7.890	7.890	7.890			
F P S			8.029	8.029	8.029		
P				8.133	8.133	8.133	
F P					8.144	8.144	
P S T					8.197	8.197	
P S					8.294	8.294	
P T						8.297	
T							8.488
Test Statistic	5.659	5.695	8.917	9.540	8.562	8.379	9.928
Adjusted Sig. (2-sided test)	.123	.258	.343	.209	.127	.218	.120

(a) Slow Sensors	Subset 1	2	3	4	5
F T	7.361				
F P T		7.755			
S		7.774	7.774		
F		7.901	7.901		
F S		7.922	7.922		
F P S T		7.934	7.934		
F P S		7.939	7.939		
F S T		7.956	7.956		
F P		8.006	8.006		
S T		8.025	8.025		
P			8.074	8.074	
P S				8.294	8.294
P T				8.304	8.304
T					8.352
P S T					8.403
Test Statistic	.	16.124	12.904	6.791	4.860
Adjusted Sig. (2-sided test)	.	.067	.184	.157	.530

(b) Normal	Subset					
Sensors	**1**	**2**	**3**	**4**	**5**	**6**
F S	7.726					
F T	7.737	7.737				
F S T	7.776	7.776				
F P T	7.778	7.778				
F		7.898	7.898			
F P S T			7.900	7.900		
S T			7.957	7.957		
S			7.976	7.976		
F P S			7.994	7.994		
F P				8.011		
T					8.133	
P T					8.143	
P S					8.215	
P S T					8.279	
P						8.476
Test Statistic	3.552	9.311	4.664	10.546	3.401	.
Adjusted Sig. (2-sided test)	.757	.092	.690	.093	.782	.

(b) Fast	Subset				
Sensors	**1**	**2**	**3**	**4**	**5**
F T	7.503				
F S	7.712	7.712			
F P S T	7.762	7.762			
S	7.772	7.772			
S T		7.834			
F S T		7.846			
F P T		7.908	7.908		
F P S		8.011	8.011	8.011	
P S T			8.127	8.127	8.127
F P			8.193	8.193	8.193
P T				8.205	8.205
P S				8.215	8.215
F				8.238	8.238
T				8.332	8.332
P					8.342
Test Statistic	8.994	12.448	10.675	13.328	10.304
Adjusted Sig. (2-sided test)	.106	.110	.050	.080	.226

APPENDIX B

Table B1: Prediction accuracy for each of the three subjects in the test cohort.

Test Subject #1

Training Set	# Samples	Hip Flex–Ext			Hip Ad–Ab			Hip Int–Ext			Hip Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.97 ± 0.02	3.4 ± 1.1	7.3 ± 2.4	0.94 ± 0.04	2.0 ± 0.6	9.4 ± 3.0	0.84 ± 0.08	2.0 ± 0.6	11.2 ± 3.4	0.92 ± 0.04	2.4 ± 0.8	9.3 ± 2.9
Synthetic	17,255	0.97 ± 0.04	0.97 ± 0.04	3.7 ± 1.7	0.94 ± 0.09	2.2 ± 0.7	10.6 ± 3.3	0.70 ± 0.15	2.7 ± 0.8	15.3 ± 4.6	0.87 ± 0.09	2.9 ± 1.1	11.2 ± 3.8
Measured + Synthetic	20,706	0.97 ± 0.01	0.97 ± 0.01	3.5 ± 0.9	0.96 ± 0.02	1.6 ± 0.4	7.4 ± 1.8	0.87 ± 0.08	2.0 ± 0.5	11.1 ± 2.8	0.94 ± 0.04	2.4 ± 0.6	8.7 ± 2.2

Training Set	# Samples	Knee Flex–Ext			Knee Ad–Ab			Knee Int–Ext			Knee Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.99 ± 0.01	2.2 ± 0.7	3.0 ± 0.9	0.49 ± 0.20	1.9 ± 0.5	14.3 ± 3.6	0.69 ± 0.11	4.6 ± 0.9	23.5 ± 4.6	0.72 ± 0.10	2.9 ± 0.7	13.6 ± 30.0
Synthetic	17,255	0.99 ± 0.01	2.3 ± 0.7	3.1 ± 0.9	0.76 ± 0.10	2.2 ± 0.6	16.9 ± 4.3	0.85 ± 0.06	2.6 ± 0.4	13.3 ± 2.3	0.87 ± 0.06	2.4 ± 0.6	11.1 ± 2.5
Measured + Synthetic	20,706	0.99 ± 0.01	1.6 ± 0.4	2.2 ± 0.6	0.91 ± 0.06	0.9 ± 0.3	7.0 ± 2.0	0.94 ± 0.04	2.1 ± 0.5	10.5 ± 2.7	0.95 ± 0.03	1.5 ± 0.4	6.6 ± 1.8

Test Subject #2

Training Set	# Samples	Hip Flex–Ext			Hip Ad–Ab			Hip Int–Ext			Hip Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.96 ± 0.05	3.6 ± 1.5	7.6 ± 3.1	0.96 ± 0.04	1.9 ± 0.6	9.1 ± 2.9	0.49 ± 0.20	5.0 ± 1.2	28.1 ± 6.6	0.80 ± 0.10	3.5 ± 1.1	15.0 ± 4.2
Synthetic	17,255	0.98 ± 0.02	2.9 ± 1.2	6.3 ± 2.6	0.96 ± 0.03	2.3 ± 0.4	10.8 ± 2.0	0.74 ± 0.15	2.7 ± 0.5	15.3 ± 2.9	0.89 ± 0.06	2.6 ± 0.7	10.8 ± 2.5
Measured + Synthetic	20,706	0.99 ± 0.01	1.9 ± 0.7	4.0 ± 1.6	0.99 ± 0.01	1.2 ± 0.4	5.7 ± 1.9	0.91 ± 0.07	2.0 ± 0.6	11.2 ± 3.3	0.96 ± 0.03	1.7 ± 0.6	7.0 ± 2.3

Training Set	# Samples	Knee Flex–Ext			Knee Ad–Ab			Knee Int–Ext			Knee Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.99 ± 0.01	2.7 ± 1.4	3.6 ± 1.8	0.80 ± 0.09	2.9 ± 0.7	22.1 ± 5.1	0.89 ± 0.07	4.4 ± 1.3	22.0 ± 6.7	0.89 ± 0.06	3.3 ± 1.1	15.9 ± 4.5
Synthetic	17,255	0.99 ± 0.01	2.0 ± 0.5	2.7 ± 0.6	0.73 ± 0.10	1.5 ± 0.4	11.5 ± 2.7	0.81 ± 0.13	3.8 ± 1.1	19.3 ± 5.7	0.84 ± 0.08	2.4 ± 0.6	11.1 ± 3.0
Measured + Synthetic	20,706	0.99 ± 0.01	1.2 ± 0.5	1.6 ± 0.7	0.91 ± 0.06	0.8 ± 0.3	6.2 ± 2.2	0.96 ± 0.03	2.1 ± 0.6	10.5 ± 3.3	0.96 ± 0.03	1.4 ± 0.5	6.1 ± 2.0

Test Subject #3

Training Set	# Samples	Hip Flex–Ext			Hip Ad–Ab			Hip Int–Ext			Hip Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.74 ± 0.08	13.3 ± 1.9	28.6 ± 4.1	0.93 ± 0.03	2.4 ± 0.6	11.5 ± 3.1	0.62 ± 0.25	5.4 ± 1.7	30.7 ± 9.4	0.76 ± 0.12	7.0 ± 1.4	23.6 ± 5.5
Synthetic	17,255	0.99 ± 0.01	1.6 ± 0.5	3.5 ± 1.0	0.95 ± 0.03	1.6 ± 0.4	7.5 ± 1.9	0.96 ± 0.02	1.5 ± 0.4	8.7 ± 2.3	0.97 ± 0.02	1.6 ± 0.4	6.5 ± 1.8
Measured + Synthetic	20,706	0.99 ± 0.01	2.4 ± 0.8	5.1 ± 1.7	0.98 ± 0.01	1.1 ± 0.5	5.3 ± 2.2	0.98 ± 0.01	1.3 ± 0.5	7.6 ± 3.0	0.98 ± 0.01	1.6 ± 0.6	6.0 ± 2.3

Training Set	# Samples	Knee Flex–Ext			Knee Ad–Ab			Knee Int–Ext			Knee Average		
		r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE	r	RMSE (°)	nRMSE
Measured	3943	0.98 ± 0.01	3.6 ± 0.8	4.9 ± 1.0	0.91 ± 0.03	1.3 ± 0.2	10.1 ± 1.8	0.72 ± 0.12	5.9 ± 0.8	30.1 ± 3.9	0.87 ± 0.05	3.6 ± 0.6	15.0 ± 2.3
Synthetic	17,255	0.99 ± 0.01	2.1 ± 0.6	2.8 ± 0.8	0.95 ± 0.02	2.2 ± 0.5	16.9 ± 4.2	0.49 ± 0.24	7.3 ± 1.1	36.7 ± 5.8	0.81 ± 0.09	3.8 ± 0.8	18.8 ± 3.6
Measured + Synthetic	20,706	0.99 ± 0.01	1.4 ± 0.4	1.8 ± 0.6	0.98 ± 0.01	0.9 ± 0.3	6.5 ± 2.6	0.90 ± 0.09	4.0 ± 1.2	20.2 ± 5.9	0.96 ± 0.04	2.1 ± 0.7	9.5 ± 3.0

APPENDIX C

Table C1: Descriptive analysis of participated patients

		Age	Height (cm)	Weight (kg)	KOOS
TKA	F	68.5±4.9	161.7±7.7	71.6±8.3	88.8±6.6
	M	68.3±2.3	176.7±6.3	91.2±25.7	92.5±8.1
OA	F	66.6±1.5	154.6±16.5	66.6±7.8	68.7±11.4
	M	65.5±9	173.9±12.2	82.4±13.3	47.9±12.6

Table C2: Top PC values and contribution on capture 75% variance of data

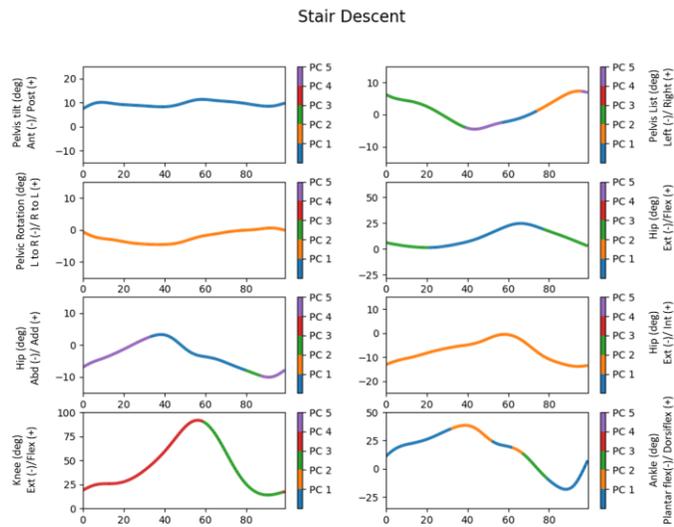| PC	Gait	Stair Ascent	Stair Descent
PC1	31%	23%	26%
PC2	23%	18%	20%
PC3	15%	17%	17%
PC4	9%	12%	10%
PC5		8%	7%

Gait

(a)

Stair Ascent

(b)

174
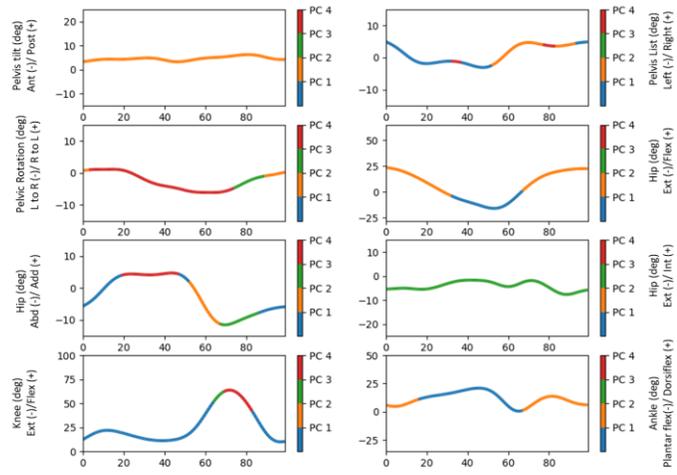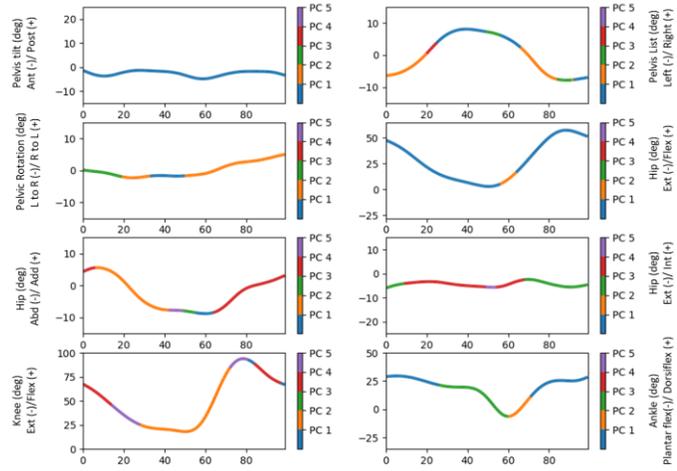
(c)

Figure C1: The dominance of PC component over the kinematic profile in a temporal manner based on absolute PC loading vector for a) gait, b) stair ascent, and c) stair descent.

Table C3: Description captured by the principal components 1 for all joints during stair descent activity.

Joint	PC	Feature	Description
Pelvis tilt	PC1	Magnitude	Magnitude of pelvic title angle throughout cycle
Pelvis list	PC1	Magnitude	Magnitude of pelvic list angle throughout cycle
Pelvis rotation	PC1	Difference	Range of motion of pelvic rotation angle throughout cycle. Difference at early stance/late swing relative to late stance/early swing.
Hip flexion	PC1	Magnitude	Magnitude of hip flexion angle throughout cycle
Hip adduction	PC1	Magnitude	Magnitude of hip adduction angle throughout cycle with more divergence at 40% of cycle.
Hip rotation	PC1	Magnitude	Magnitude of hip internal rotation angle throughout cycle
Knee flexion	PC1	Difference	Range of motion of knee flexion angle throughout cycle. Difference at swing phase relative to early stance phase
Ankle dorsiflexion	PC1	Magnitude	Magnitude of ankle dorsiflexion throughout cycle

Table C4: Description captured by the principal components 2 and 3 for all joints during gait activity.

Joint	PC	Feature	Description
Pelvis tilt	PC2	Magnitude	Magnitude of pelvic title angle throughout cycle
	PC3	NA	NA
Pelvis list	PC2	Magnitude	Magnitude of pelvis list angle at swing phase
	PC3	Difference	Range of motion of pelvis tilt angle throughout cycle. Difference at swing phase relative to stance phase.
Pelvis rotation	PC2	Magnitude	Magnitude of pelvic rotation angle throughout cycle
	PC3	Magnitude	Magnitude of pelvic rotation angle throughout cycle
Hip flexion	PC2	Magnitude	Magnitude of hip flexion angle throughout cycle
	PC3	NA	NA
Hip adduction	PC2	Difference	Range of motion of hip adduction angle throughout cycle. Difference at early swing phase relative to stance phase.
	PC3	Difference	Range of motion of hip adduction angle throughout cycle. Difference at swing phase relative to early stance/stance phases.
Hip rotation	PC2	NA	NA
	PC3	Magnitude	Magnitude of hip internal rotation angle throughout cycle
Knee flexion	PC2	Difference/ Magnitude	Range of motion of knee flexion angle throughout cycle. Difference at swing phase relative to stance phase with more dominance at stance phase
	PC3	Difference/ Magnitude	Range of motion of hip adduction angle throughout cycle. Difference at swing phase relative to stance phase with more dominance at swing phase
Ankle dorsiflexion	PC2	Magnitude	Magnitude of ankle dorsiflexion throughout cycle
	PC3	NA	NA

Table C5: Description captured by the principal component 4 for all joints during stair ascent activity.

Joint	PC	Feature	Description
Pelvis tilt	PC4	Magnitude	Magnitude of pelvic title angle throughout cycle
Pelvis list	PC4	Difference	Range of motion of pelvis list angle throughout cycle. Difference at swing phase relative to stance phase
Pelvis rotation	PC4	Magnitude	Magnitude of pelvis rotation angle throughout cycle
Hip flexion	PC4	Difference	Range of motion of hip flexion angle throughout cycle. Difference at swing phase relative to stance/end phase of the cycle.
Hip adduction	PC4	Difference	Range of motion of hip adduction angle throughout cycle. Difference at stance phase relative to early stance/swing phase of the cycle.
Hip rotation	PC4	Magnitude	Magnitude of hip internal rotation angle throughout cycle

176

| Knee flexion | PC4 | Difference | Range of motion of knee flexion angle throughout cycle. Difference at swing phase relative to stance/late swing phases. |
| Ankle dorsiflexion | PC4 | Magnitude | Magnitude of ankle dorsiflexion throughout cycle |

Table C6: Mean and covariate matrix of OA and TKA groups across activities of Gait, Stair Ascent, Stair Descent as well as the mean and standard deviation used for standardization.

Gait	PC1	PC2	PC3	PC4	Gait
OA Mean	0.3990	-0.4574	-0.4506	0.2111	OA Mean
OA Cov1	0.8702	0.3118	0.2944	0.6027	OA Cov1
OA Cov2	0.3118	0.9059	-0.1842	0.0842	OA Cov2
OA Cov3	0.2944	-0.1842	0.7191	0.3657	OA Cov3
OA Cov4	0.6027	0.0842	0.3657	1.4134	OA Cov4
TKA Mean	-0.3242	0.3716	0.3661	-0.1715	TKA Mean
TKA Cov1	0.9203	0.0373	0.0469	-0.6145	TKA Cov1
TKA Cov2	0.0373	0.8133	-0.1764	0.0843	TKA Cov2
TKA Cov3	0.0469	-0.1764	0.9724	-0.1431	TKA Cov3
TKA Cov4	-0.6145	0.0843	-0.1431	0.6660	TKA Cov4
Original Mean	1.2251e-15	2.0000e-09	-1.8621e-09	5.5172e-10	Original Mean
Original Std	96.8340	84.2592	67.2238	51.9743	Original Std

Stair Ascent	PC1	PC2	PC3	PC4	PC5
OA Mean	0.3415	0.3976	0.1130	-0.4129	0.0580
OA Cov1	1.1446	-0.3856	-0.0227	0.3766	-0.1378
OA Cov2	-0.3856	1.5234	-0.2472	0.4171	0.3224
OA Cov3	-0.0227	-0.2472	0.9731	-0.0717	0.3939
OA Cov4	0.3766	0.4171	-0.0717	0.9457	-0.2810
OA Cov5	-0.1378	0.3224	0.3939	-0.2810	1.0086
TKA Mean	-0.2774	-0.3231	-0.0918	0.3354	-0.0471
TKA Cov1	0.7679	0.0952	-0.0424	-0.0798	0.0792
TKA Cov2	0.0952	0.3995	0.1272	-0.0758	-0.2941
TKA Cov3	-0.0424	0.1272	1.0681	0.1306	-0.3254
TKA Cov4	-0.0798	-0.0758	0.1306	0.8424	0.2624
TKA Cov5	0.0792	-0.2941	-0.3254	0.2624	1.0545
Original Mean	2.1034e-09	1.8276e-09	-1.0345e-10	-3.2759e-09	1.3103e-09
Original Std	96.3232	85.3912	81.8647	69.2004	57.8539

Stair Descent	PC1	PC2	PC3	PC4	PC5
OA Mean	-0.6882	-0.0963	-0.3066	-0.0913	0.1087
OA Cov1	0.5895	-0.1069	-0.2519	0.0658	0.1091
OA Cov2	-0.1069	1.1021	0.3564	0.0579	0.0246
OA Cov3	-0.2519	0.3564	0.9461	-0.4210	-0.5535
OA Cov4	0.0658	0.0579	-0.4210	1.6581	-0.0202
OA Cov5	0.1091	0.0246	-0.5535	-0.0202	1.3512
TKA Mean	0.5591	0.0782	0.2491	0.0741	-0.0883
TKA Cov1	0.6511	-0.0186	-0.1299	-0.1513	0.0302
TKA Cov2	-0.0186	0.9704	-0.3315	-0.0601	-0.0032
TKA Cov3	-0.1299	-0.3315	0.9621	0.2928	0.4952
TKA Cov4	-0.1513	-0.0601	0.2928	0.5271	0.0317
TKA Cov5	0.0302	-0.0032	0.4952	0.0317	0.7672
Original Mean	-6.8968e-11	-1.3793e-10	-1.6207e-09	-2.8965e-09	1.5517e-09
Original Std	103.2869	92.4932	85.9263	64.3180	53.5996
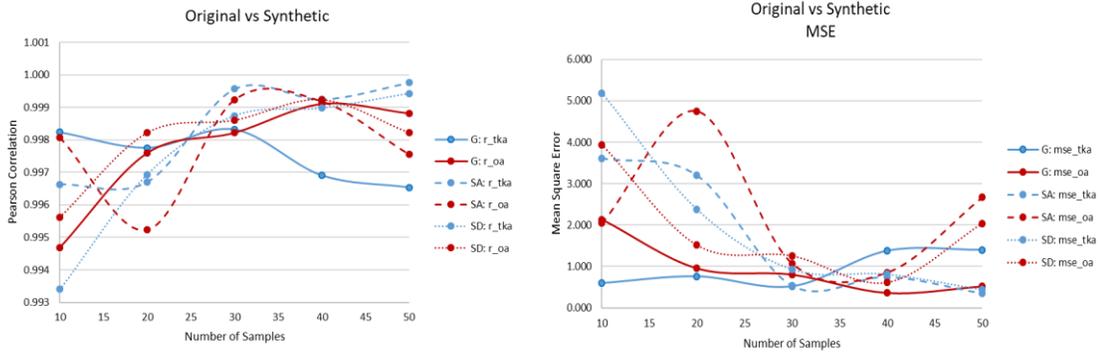
(a)   (b)

Figure C2: The Pearson correlation (a) and mean square error (b) between average original and synthetic kinematic at 10, 20, 30, 40, and 50 samples for Gait (G), Stair Ascent (SA), and Stair Descent (SD)
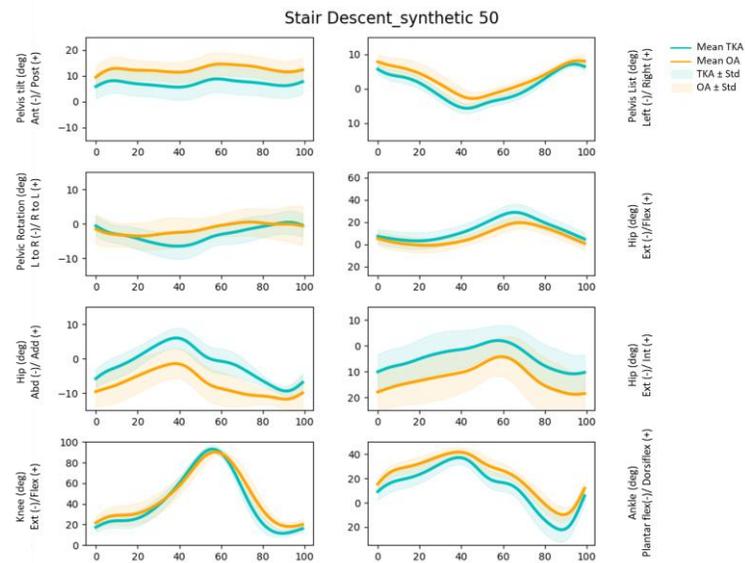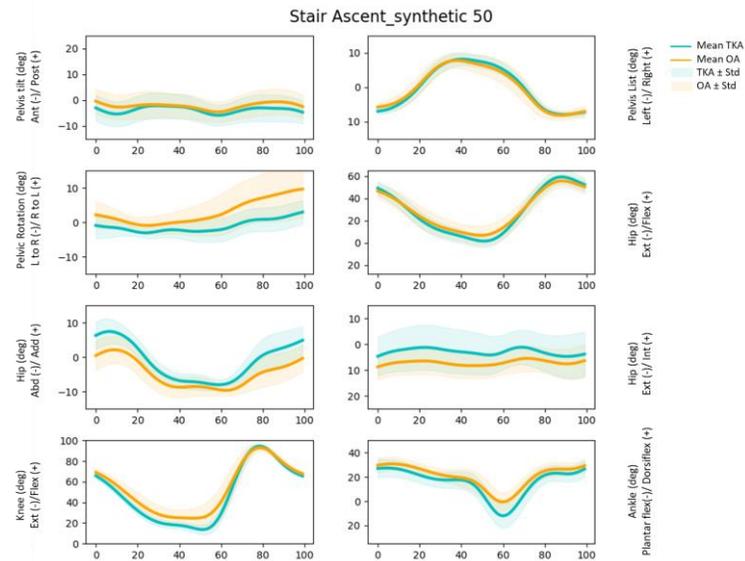

(a)

(b)

(c)
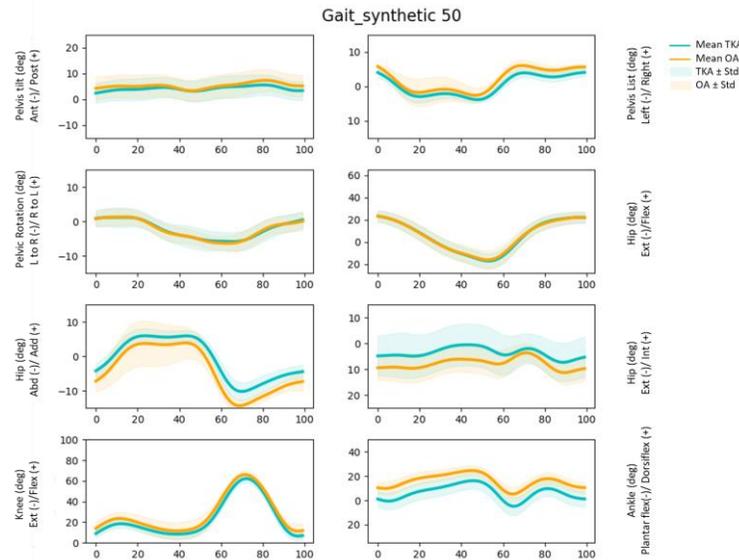
Figure C3: The kinematic mean and standardization of 50 synthetic kinematic data for
a) gait, b) stair ascent, and c) stair descent.